Probabilités

Préparation à l'agrégation interne

Djalil Chafaï

Copyright © 2012 (sauf image ci-dessus)

Livre électronique enregistré à l'AFNIL
http://djalil.chafai.net/enseignement.html
ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Table des matières

0 Avant propos **5**

1 Modélisation d'une expérience **9**

2 Espace probabilisé **13**
2.1 Conditionnement 18
2.2 Indépendance 19

3 Variables aléatoires réelles **23**
3.1 Fonction de répartition et loi 23
3.2 Espérance 26
3.3 Espérance des v.a.r. discrètes 28
3.4 Espérance des v.a.r. à densité 31
3.5 Variance 32
3.6 Inégalités de Markov et de Bienaymé-Tchebychev 33

4 Vecteurs aléatoires **35**
4.1 Indépendance et covariance 37
4.2 Matrice de covariance 40
4.3 Fonctions génératrices 42

5 Théorèmes limites **43**
5.1 Loi des grands nombres 43
5.2 Théorème de la limite centrale 48
5.3 Approximation de la loi Binomiale par la loi normale 50
5.4 Distance en variation 53
5.5 Approximation de la loi Binomiale par la loi de Poisson 55

A Compléments **59**
A.1 Lois exponentielles 59
A.2 Jeu de pile ou face 64
A.3 Collectionneur de coupons 66
A.4 Marche aléatoire simple et ruine du joueur 69
A.5 Fonctions caractéristiques et vecteurs gaussiens 71
A.6 Extrêmes 77

B Leçons d'oral **81**

Copyrighted material

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Chapitre 0

Avant propos

Ces notes de cours couvrent l'essentiel des notions de probabilités au programme de l'agrégation interne de mathématiques. **Elles ne constituent en aucun cas des modèles de leçons d'oral**. L'image de couverture provient d'Internet, son propriétaire est inconnu.

Ce livre électronique possède un ISBN fourni par l'AFNIL. Il est donc en principe utilisable le jour de l'oral par les candidats à partir de la session 2013. Si vous l'appréciez, vous pouvez manifester votre gratitude en faisant un don directement à l'auteur sur `http://djalil.chafai.net/enseignement.html`. C'est vous qui choisissez le montant, tandis que pour un livre classique, le prix est fixe et l'essentiel revient à l'éditeur. Cette rémunération directe de l'auteur est dans le même esprit que la rémunération directe des agriculteurs et artistes. Elle court-circuite les marchands, qui ne produisent rien. Mais vous pouvez aussi choisir de l'utiliser sans payer, car le savoir doit rester librement accessible.

Voici un extrait du programme 2011 (section 13 et parties de la section 9). On notera l'absence de l'intégrale de Lebesgue et des fonctions caractéristiques.

9 Analyse réelle et complexe

...

9.6 Intégrale d'une fonction continue par morceaux sur un segment

Définition de l'intégrale de Riemann, linéarité, positivité, inégalité de la moyenne, relation de Chasles. Inégalité de Cauchy-Schwarz.

Primitive d'une fonction continue sur un intervalle. Intégration par parties, changement de variable, calculs de primitives et d'intégrales.

Convergences en moyenne et en moyenne quadratique pour les suites de fonctions. Comparaison avec la convergence uniforme.

...

9.8 Intégration sur un intervalle quelconque

Les fonctions considérées dans ce paragraphe sont supposées continues par morceaux sur l'intervalle I de définition, c'est-à-dire continues par morceaux sur tout segment contenu dans I.

Intégrale d'une fonction positive (comme borne supérieure, éventuellement infinie, des intégrales sur les segments inclus dans I). Emploi des relations de comparaison.

Une fonction définie sur I à valeurs complexes est dite intégrable si l'intégrale de son module est finie.

Les trois théorèmes suivants sont admis :

Théorème de convergence monotone : Soit (f_n) une suite croissante de fonctions intégrables, convergeant simplement sur I vers une fonction f continue par morceaux sur I. Alors f est intégrable sur I si, et seulement si, la suite des intégrales des f_n est majorée ; en ce cas, l'intégrale de f est la limite de celles des f_n .

Théorème de convergence dominée : Soit (f_n) une suite de fonctions à valeurs complexes convergeant simplement sur I vers une fonction f continue par morceaux sur I. Si la suite des modules des f_n est majorée par une fonction g intégrable sur I, alors f est intégrable sur I et son intégrale est la limite de celles des f_n.

Théorème d'intégration terme à terme : Soit une suite (u_n) de fonctions à valeurs complexes, intégrables sur I, telle que la série $\sum_n u_n$ converge simplement vers une fonction S continue par morceaux sur I, et telle que la série $\sum_n \int_I |u_n|$ converge. Alors S est intégrable sur I et on a $\int_I S = \sum_n \int_I u_n$.

9.9 Intégrales impropres

Intégrales convergentes, divergentes ; critère de Cauchy.

Copyrighted material

Convergence absolue, lien avec l'intégrabilité. Emploi des relations de comparaison, de l'intégration par parties pour l'étude de la convergence. Intégration de relations de prépondérance et d'équivalence.
Pour une fonction f définie sur $[a,+\infty[$ et à valeurs positives, comparaison entre la convergence de la série de terme général $f(n)$ $(n \geq a)$ et l'intégrabilité de f sur $[a,+\infty[$ (méthode des rectangles). Si f est décroissante et positive sur $[0,+\infty[$, alors la série de terme général $f(n) - \int_{[n,n+1]} f(t)\,dt$ converge.

13. Calcul intégral et probabilités

13.1 Intégrales multiples *Tous les théorèmes de ce paragraphe sont admis.*
Intégrales curvilignes, longueur d'un arc de courbe, travail d'une force.
Formule de Fubini et définition de l'intégrale double d'une fonction continue sur un rectangle $[a,b] \times [c,d]$. Adaptation à l'intégrale triple.
Théorème de Fubini-Tonelli : Si f est une fonction de deux variables continue positive sur un rectangle borné ou non, on peut intervertir l'ordre des intégrations ; lorsque la valeur commune de ces intégrales est finie, f est dite intégrable et son intégrale double est cette valeur commune.
Si f est une fonction complexe de deux variables continue sur un rectangle borné ou non, on dit que f est intégrable si son module est intégrable. Dans ce cas, on peut intervertir l'ordre des intégrations et l'intégrale de f est la valeur commune des deux intégrales superposées.
Extension des résultats précédents au cas de fonctions de plusieurs variables.
Extension au cas du produit d'une fonction de plusieurs variables continue positive par une fonction indicatrice d'un ensemble «géométriquement simple». Linéarité et additivité relativement à la fonction et relativement aux ensembles.
Applications à des calculs d'intégrales.
Théorème du changement de variables ; passage en coordonnées polaires.
Exemples de calculs d'aires planes et de volumes.

13.2 Modélisation d'une expérience aléatoire
Espace Ω des épreuves (ou des évènements élémentaires)
tribu (ou σ-algèbre) $\mathcal{F}$ des évènements ;
mesure de probabilité $\mathbb{P}$ sur cette tribu.
Étude d'exemples dans le cas où Ω est fini ou infini dénombrable.

13.3 Espace probabilisé
Propriétés d'une probabilité.
Probabilité conditionnelle $\mathbb{P}_B(A)$ de A sachant B si $\mathbb{P}(B)$ est non nul.
Formule des probabilités composées (ou totales) et formule de Bayes.
Indépendance d'un ensemble fini d'évènements.

13.4 Variables aléatoires réelles
Étant donné un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, on appelle variable aléatoire réelle (v.a.r. en abrégé) toute application X de Ω dans $\mathbb{R}$ telle que l'image réciproque $X^{-1}(I)$ de tout intervalle I de $\mathbb{R}$ appartienne à la tribu $\mathcal{F}$. On admettra que la somme, ou le produit, de v.a.r. est une v.a.r.. On se bornera à l'étude des deux familles suivantes de v.a.r. :

13.4.1 Variables aléatoires réelles discrètes
Une v.a.r. est dite discrète si elle prend un nombre fini ou infini dénombrable de valeurs. Loi et fonction de répartition d'une v.a.r. discrète. Moments d'une v.a.r. discrète : espérance, variance et écart type. Espérance d'une somme de v.a.r. discrètes. Fonction génératrice d'une v.a.r. à valeurs dans $\mathbb{N}$. Lois discrètes usuelles : loi hypergéométrique, loi de Bernoulli, loi binomiale, loi géométrique et loi de Poisson.

13.4.2 Variables aléatoires réelles possédant une loi avec densité
On appelle densité de probabilité sur $\mathbb{R}$ toute fonction de $\mathbb{R}$ dans $\mathbb{R}_+$ intégrable sur $\mathbb{R}$ et d'intégrale égale à 1 (on se limitera à la notion d'intégrale définie dans le paragraphe 9.8). Soit f une densité de probabilité sur $\mathbb{R}$. On dit qu'une v.a.r. X possède la loi de densité f si, pour tout intervalle I de $\mathbb{R}$, $\mathbb{P}(\{X \in I\}) = \int_I f(x)\,dx$.
Fonction de répartition et moments ; espérance, variance et écart type d'une v.a.r. possédant une loi avec densité. Espérance d'une somme de v.a.r. possédant une densité (résultat admis). Lois usuelles possédant une densité : loi uniforme sur un intervalle borné ; loi exponentielle ; loi de Cauchy ; loi normale.
On admettra le résultat suivant (théorème de transfert) : si X est une v.a.r. de loi de densité f et si Φ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$ continue par morceaux sur tout segment et telle que la fonction $|\Phi|f$ soit intégrable sur $\mathbb{R}$, alors $\Phi(X)$ est une v.a.r. dont l'espérance est donnée par : $\mathbb{E}(\Phi(X)) = \int_{\mathbb{R}} \Phi(x) f(x)\,dx$.

13.5 Vecteurs aléatoires
On dira qu'une application $X = (X_1, \ldots, X_p)$ de Ω dans $\mathbb{R}^p$ est un vecteur aléatoire si chacune de ses composantes est une v.a.r. On se limitera aux deux cas suivants :

13.5.1 Vecteurs aléatoires discrets
Un vecteur aléatoire $X = (X_1, \ldots, X_p)$ de Ω dans $\mathbb{R}^p$ est dit discret si chacune de ses composantes est une v.a.r. discrète. Loi d'un vecteur aléatoire X. Indépendance de p v.a.r. discrètes. Covariance et coefficient de corrélation d'un couple de v.a.r. discrètes. Espérance et variance d'une somme de p v.a.r. discrètes indépendantes.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

13.5.2 Vecteurs aléatoires possédant une loi avec densité

On appelle densité de probabilité sur $\mathbb{R}^p$ toute fonction f de $\mathbb{R}^p$ dans $\mathbb{R}^+$, intégrable sur $\mathbb{R}^p$ et d'intégrale égale à 1 (on se limitera à la notion d'intégrale définie dans le paragraphe 13.1). Soit f une densité de probabilité sur $\mathbb{R}^p$. On dit qu'un vecteur aléatoire $X = (X_1, \ldots, X_p)$ possède la loi de densité f si on a, pour tous intervalles $I_1, \ldots, I_p$ de $\mathbb{R}$,

$$\mathbb{P}(\{X_1 \in I_1\} \cap \cdots \cap \{X_p \in I_p\}) = \int_{I_1} \cdots \int_{I_p} f(x_1, \ldots, x_p)\, dx_1 \cdots dx_p.$$

Soit $X = (X_1, \ldots, X_p)$ un vecteur aléatoire de loi de densité f. Soit Ψ un produit d'une fonction continue de $\mathbb{R}^p$ dans $\mathbb{R}$ par une fonction indicatrice d'un domaine « géométriquement simple » de $\mathbb{R}^p$ et telle que la fonction $|\Psi| f$ soit intégrable sur $\mathbb{R}^p$. On admettra que $\Psi(X)$ est une v.a.r. dont l'espérance est donnée par :

$$\mathbb{E}(\Psi(X)) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \Psi(x_1, \ldots, x_p) f(x_1, \ldots, x_p)\, dx_1 \cdots dx_p$$

Indépendance de p v.a.r. possédant une loi avec densité. Covariance et coefficient de corrélation d'un couple de v.a.r. possédant une loi avec densité. Espérance et variance d'une somme de p v.a.r. indépendantes et possédant une loi avec densité. Application aux loi normales.

13.6 Théorèmes limites

Suites de v.a.r. indépendantes.

Inégalité de Bienaymé-Tchebychev et loi faible des grands nombres.

Lemme de Borel-Cantelli.

Les résultats suivants sont admis : loi forte des grands nombres pour une suite de v.a.r. indépendantes équidistribuées possédant une espérance. Théorème de la limite centrale pour une suite de v.a.r. indépendantes équidistribuées et de variance finie.

Approximations de la loi binomiale par la loi de Poisson et par la loi normale.

Voici enfin une bibliographie probabiliste tirée de la liste d'ouvrages disponibles le jour de l'oral d'après le rapport 2010 du jury :

- ⋆ Barbe et Ledoux, *Probabilité*
- Bercu et Chafaï, *Modélisation stochastique et simulation*
- Bon, *Fiabilité des systèmes*
- Brémaud, *Introduction aux probabilités*
- Combrouze, *Probabilités et statistique*
- ⋆ Cottrell et Genon-Catalot et Duhamel et Meyre, *Exercices de probabilités*
- Dacunha-Castelle et Duflo, *Probabilités et statistique 1*
- Dacunha-Castelle et Duflo, *Exercices de probabilités et statistique 1*
- Dacunha-Castelle, Revuz, et Schreiber, *Recueil de problèmes de calcul des probabilités*
- ⋆ Dantzer, *Maths pour l'agrég. interne, Analyse et probas. Cours et exos corrigés*
- Dembo et Zeitouni, *Large deviations techniques and applications*
- Feller, *An introduction to probability theory & its applications*
- Grimmett et Welsh, *Probability (an introduction)*
- Hammad, *Cours de probabilités*
- Hammad et Taranco, *Exercices de probabilités*
- Hennequin et Tortrat, *Théorie des probabilités et quelques applications*
- Isaac, *Une initiation aux probabilités*
- Métivier, *Notions fondamentales de la théorie des probabilités*
- Métivier, *Probabilités : dix leçons d'introduction*
- Neveu, *Bases mathématiques du calcul des probabilités*
- Norris, *Markov chains*
- ⋆ Ouvrad, *Probabilités 1 (Capes et agrégation) et 2 (maîtrise et agrégation)*
- Pages et Bouzitat, *En passant par le hasard, les probabilités de tous les jours*
- Rao, *Linear statistical inference and its application*
- Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*
- Robert, *Contes et décomptes de la statistique - Une initiation par l'exemple*
- Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*
- Tenenbaum, *Exercices corrigés de théorie analytique et probabiliste des nombres*
- Tortrat, *Calcul des probabilités et introduction aux processus aléatoires*
- Vauquois, *Outils mathématiques. Probabilités*

Copyright © Djalil Chafaï, 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

– Warusfel, Attali, Collet, Gautier, et Nicolas, *Mathématiques, probabilités*

Les références étoilées ont été appréciées par les agrégatifs des années précédentes.

Bonne lecture !

Djalil Chafaï
Marne-la-Vallée
Hiver 2012

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Chapitre 1

Modélisation d'une expérience

Ensembles	Évènements
Ω	certitude
$\varnothing$	impossible
$A \subset B$	A implique B
$A \cap B$	A et B c'est-à-dire à la fois A et B
$A \cup B$	A ou B c'est-à-dire au moins A ou B
$A \cap B = \varnothing$ (A et B disjoints)	A et B sont incompatibles
A^c (on note aussi $\overline{A}$)	contraire de A

Table 1.1 – Approche ensembliste des probabilités. On écrit *événement* ou *évènement*.

On modélise une expérience aléatoire en introduisant l'ensemble Ω encodant tous les résultats possibles de l'expérience, appelé *univers*. Voici quelques exemples concrets :

- un lancer de pièce de monnaie (pile ou face) : $\Omega = \{0, 1\}$
- deux lancers consécutifs : $\Omega = \{0, 1\} \times \{0, 1\} = \{0, 1\}^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$
- n lancers de pièce de monnaie consécutifs : $\Omega = \{0, 1\}^n$
- une infinité de lancers de pièce de monnaie consécutifs : $\Omega = \{0, 1\}^{\mathbb{N}}$
- un jet de dé à 6 faces : $\Omega = \{1, 2, 3, 4, 5, 6\}$
- un jet de deux dés à 6 faces de $\neq$ couleurs : $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$
- nombre d'étoiles observables la nuit : $\Omega = \mathbb{N}$
- cote de popularité de Nicolas : $\Omega = [0, 100]$
- durée de vie d'une ampoule : $\Omega = \mathbb{R}_+$
- poids d'un être humain : $\Omega = [0, 500]$
- point d'impact au jeu de fléchettes : $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$
- température Celcius de la soupe du soir : $[-273, +\infty]$
- position du moustique : $\Omega = \mathbb{R}^3$
- état d'un jeu de n cartes après battage : $\Omega = \Sigma_n$ (groupe symétrique)
- cours d'une action en bourse sur la période T : $\Omega = \mathcal{C}(T, \mathbb{R}_+)$

Il règne là un arbitraire typique de l'étape de modélisation, qu'on ne peut pas évacuer complètement. On dit souvent à ce propos que *tous les modèles sont faux, mais que certains sont plus utiles que d'autres*. Cet arbitraire de la modélisation n'est pas spécifique aux probabilités. Il se trouve simplement que les modélisations liées à l'analyse (analyse numérique par exemple), à l'algèbre (cryptographie par exemple), ou à la géométrie (cartographie par exemple) ne figurent pas au programme.

On voit bien que Ω peut être fini, ou infini dénombrable, ou même infini non dénom-

Copyrighted material

brable. Dans le cas du jet de dé à 6 faces, le sous ensemble $\{2,4,6\} \subset \Omega$ s'identifie à l'évènement *le résultat est pair*. L'évènement *impossible* s'identifie à l'ensemble $\varnothing$ tandis que l'évènement *certain* s'identifie à l'ensemble Ω tout entier. Plus généralement, les évènements d'intérêt s'identifient à une collection de parties $\mathcal{F} \subset \mathcal{P}(\Omega)$. Deux évènements sont *incompatibles* lorsqu'ils sont disjoints en tant qu'ensembles. Il est naturel d'imposer à $\mathcal{F}$ quelques propriétés de stabilité, qui en font une *tribu* (on dit parfois également σ-*algèbre*).

Définition 1.1 (Tribu). *On dit qu'une collection $\mathcal{F} \subset \mathcal{P}(\Omega)$ constitue une* tribu *lorsque*

1. $\Omega \in \mathcal{F}$
2. ***complémentaire*** *: pour tout $A \in \mathcal{F}$ on a $A^c \in \mathcal{F}$*
3. $\cup$ ***dénombrable*** *: pour toute suite (A_n) d'éléments de $\mathcal{F}$ on a $\cup_n A_n \in \mathcal{F}$*

Rappelons que si $(A_i)_{i\in I}$ est une famille d'évènements et B un évènement alors on a

$$\text{Formules de Morgan} \quad \begin{cases} (\cup_{i\in I} A_i)^c &= \cap_{i\in I} A_i^c \\ (\cap_{i\in I} A_i)^c &= \cup_{i\in I} A_i^c \end{cases}$$

$$\text{Distribuvité} \quad \begin{cases} B \cap (\cup_{i\in I} A_i) &= \cup_{i\in I}(B \cap A_i) \\ B \cup (\cap_{i\in I} A_i) &= \cap_{i\in I}(B \cup A_i) \end{cases}$$

Proposition 1.2 (Propriétés des tribus). *Si $\mathcal{F}$ est une tribu sur Ω alors...*

1. $\varnothing \in \Omega$
2. $\cap$ ***dénombrable*** *: pour toute suite (A_n) d'éléments de $\mathcal{F}$ on a $\cap_n A_n \in \mathcal{F}$*
3. **limite inférieure *et* limite supérieure** *: si $(A_n)_{n\geq 1}$ est une suite d'éléments de $\mathcal{F}$ alors les évènements suivants appartiennent également à $\mathcal{F}$:*

$$\underline{\lim}\, A_n = \bigcup_n \bigcap_{m\geq n} A_m = \{\omega \in \Omega \text{ t.q. } \omega \in A_n \text{ à partir d'un certain rang sur } n\}$$

et

$$\overline{\lim}\, A_n = \bigcap_n \bigcup_{m\geq n} A_m = \{\omega \in \Omega \text{ t.q. } \omega \in A_n \text{ pour une infinité de valeurs de } n\}.$$

et de plus

$$(\underline{\lim}\, A_n)^c = \overline{\lim}\, A_n^c \quad \text{et} \quad (\overline{\lim}\, A_n)^c = \underline{\lim}\, A_n^c.$$

Exemple 1.3 (Exemples de tribus).

1. *la* tribu triviale $\{\varnothing, \Omega\}$, *et la* tribu grossière $\mathcal{P}(\Omega)$
2. *si $A \subset \Omega$ alors $\{\varnothing, A, A^c, \Omega\}$ est une tribu*
3. *si $(\mathcal{F}_i)_{i\in I}$ est une famille quelconque de tribus sur Ω alors $\cap_{i\in I}\mathcal{F}_i$ est une tribu sur Ω. La* ***tribu engendrée*** *par $\mathcal{A} \subset \mathcal{P}(\Omega)$, $\mathcal{A} \neq \varnothing$, est l'intersection de toutes les tribus contenant $\mathcal{A}$. C'est la plus petite (pour l'inclusion) tribu contenant $\mathcal{A}$*
4. *si Ω est dénombrable (fini ou infini) on choisit en général la tribu engendrée par les singletons, qui n'est rien d'autre que $\mathcal{P}(\Omega)$. Cette tribu est également appelée tribu atomique car un singleton $\{\omega\}$ avec $\omega \in \Omega$ est appelé* atome.
5. *si $\Omega = \mathbb{R}^d$ on choisit en général la tribu borélienne $\mathcal{B}$ qui est la tribu engendrée par les pavés de $\mathbb{R}^d$ (i.e. par les intervalles lorsque $d = 1$).*

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

6. *si $\Omega = \Omega_1 \times \Omega_2$ et si $\mathcal{F}_1$ et $\mathcal{F}_2$ sont des tribus sur Ω_1 et Ω_2 respectivement alors on choisit en général la* ***tribu produit*** *sur Ω, notée*

$$\mathcal{F}_1 \otimes \mathcal{F}_2$$

engendrée par les produits $A_1 \times A_2$ où $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$. On admet que la tribu borélienne sur $\mathbb{R}^d$ est la tribu produit des tribus boréliennes.

7. *si $\Omega = \{0,1\}^{\mathbb{N}}$ (pas dénombrable : argument diagonal de Cantor) on choisit la* ***tribu cylindrique*** *engendrée par les* ***cylindres***

$$A_0 \times A_1 \times A_2 \times \cdots$$

où $A_0, A_1, A_2, \ldots \in \{0,1\}$ et $A_n = \{0,1\}$ à partir d'un certain rang sur n. Cela permet de modéliser un jeu de pile ou face infini. Plus généralement, si $\mathcal{F}'$ est une tribu sur Ω' alors on équipe $\Omega = \Omega'^{\mathbb{N}}$ de la tribu engendrée par les cylindres $A_0 \times A_1 \times A_2 \times \cdots$ où $A_0, A_1, A_2, \ldots \in \mathcal{F}'$ et $A_n = \Omega'$ à partir d'un certain rang sur n.

La *fonction indicatrice* d'un évènement $A \subset \Omega$ est la fonction booléenne

$$\mathbf{1}_A : \omega \in \Omega \mapsto \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Les fonctions indicatrices servent à compter : la somme $\sum_n \mathbf{1}_{A_n}$ est égale au nombre d'événement A_n qui ont lieu. Voici d'autres exemples :

Théorème 1.4 (Indicatrices). *Si $A, B \in \mathcal{F}$ et $(A_n)_{n\geq 1}$ sont dans la tribu $\mathcal{F}$ alors...*

1. $\mathbf{1}_{A\cap B} = \mathbf{1}_A \mathbf{1}_B$ *et* $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}$
2. $\underline{\lim}\, A_n = \{\sum_n \mathbf{1}_{A_n^c} < \infty\}$ *et* $\overline{\lim}\, A_n = \{\sum_n \mathbf{1}_{A_n} = \infty\}$
3. $\mathbf{1}_{\underline{\lim}\, A_n} = \underline{\lim}\, \mathbf{1}_{A_n}$ *et* $\mathbf{1}_{\overline{\lim}\, A_n} = \overline{\lim}\, \mathbf{1}_{A_n}$

Notons que $\mathbf{1}_A^2 = \mathbf{1}_A$ pour tout $A \in \mathcal{F}$. Les fonctions indicatrices vont jouer un rôle important après l'introduction des notions de $\mathbb{P}$ et $\mathbb{E}$ car $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Chapitre 2

Espace probabilisé

Les mesures de probabilité permettent de comparer l'importance des évènements. On peut les interpréter comme un cardinal relatif ou une surface relative par rapport au tout.

Définition 2.1 (Probabilité). *Soit $\mathcal{F}$ une tribu sur un univers Ω. Une* ***mesure de probabilité*** *ou* ***loi de probabilité*** *sur $(\Omega, \mathcal{F})$ est une application $\mathbb{P} : \mathcal{F} \to [0,1]$ vérifiant*

1. ***Normalisation :*** $\mathbb{P}(\Omega) = 1$
2. ***Σ-additivité :*** *si $(A_n)_{n\geq 1}$ est une suite d'événements deux à deux disjoints*[1] *alors*

$$\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$$

On dit que le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est un ***espace probabilisé****.*

Proposition 2.2 (Propriétés immédiates). *Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé alors*

1. $\mathbb{P}(\varnothing) = 0$
2. *pour tout $A \in \mathcal{F}$ on a $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*
3. *pour tous $A, B \in \mathcal{F}$ on a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$*
4. *pour tous $A, B \in \mathcal{F}$ si $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$*
5. *pour tous $A, B \in \mathcal{F}$ on a $\mathbb{P}(A)\mathbb{P}(B) \leq \min(\mathbb{P}(A), \mathbb{P}(B))$*
6. *pour tous $A, B \in \mathcal{F}$ avec $B \subset A$ on a $\mathbb{P}(A) - \mathbb{P}(B) = \mathbb{P}(A \cap B^c)$*
7. ***probabilités totales*** *: si (A_n) est une partition de Ω alors $\sum_n \mathbb{P}(A_n) = 1$*

Démonstration. Pour la troisième propriété, on utilise les partitions

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \quad \text{et} \quad A = (A \setminus B) \cup (A \cap B).$$

□

Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé avec Ω **fini**. On dit qu'on a **équiprobabilité** lorsque $\mathbb{P}$ est la *mesure de probabilité uniforme sur* Ω, c'est-à-dire que pour tout $\omega \in \Omega$,

$$\mathbb{P}(\{\omega\}) = \frac{1}{\mathrm{Card}(\Omega)} \quad \text{pour tout } \omega \in \Omega.$$

Tous les atomes de Ω ont la même probabilité. Dans ce cas, on a pour tout $A \subset \Omega$,

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}.$$

1. Cela signifie que $A_i \cap A_j = \varnothing$ si $i \neq j$. À ne pas confondre avec $\cap_n A_n = \varnothing$.

Copyrighted material

Il s'agit de la fameuse formule « cas favorables sur cas totaux ». Elle n'est **valable qu'en cas d'équiprobabilité**, et réduit le calcul des probabilités à du dénombrement. Un lancer de pile ou face avec une pièce équilibrée se modélise avec la mesure de probabilité uniforme sur $\{0,1\}$, qui affecte la probabilité $\frac{1}{2}$ aux 2 atomes $\{0\}$ et $\{1\}$. Un jet de dé équilibré à six faces se modélise avec la mesure de probabilité sur $\{1,2,3,4,5,6\}$ qui affecte la probabilité $\frac{1}{6}$ aux 6 atomes $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$.

Remarque 2.3 (Équiprobabilité sur les univers infinis)**.** *Il ne peut pas y avoir d'équiprobabilité lorsque* $(\Omega,\mathcal{F}) = (\mathbb{N},\mathcal{P}(\mathbb{N}))$ *car la masse d'un atome serait nulle. Il n'y a donc pas de mesure de probabilité uniforme sur les ensembles infinis dénombrables. En revanche, si* Ω *est un pavé de* $\mathbb{R}^d$ *équipé de la tribu borélienne, alors la mesure de Lebesgue normalisée joue le rôle de modèle d'équiprobabilité, à condition de remplacer le cardinal par le volume.*

En pratique, les modèles équiprobables sont les plus naturels, et constituent le socle sur lequel tous les autres sont construits. Certaines de ces constructions sont abordées dans la suite. Les calculs de cardinaux dans les modèles d'équiprobabilité nécessitent bien souvent des formules combinatoires. Rappel : on considère une urne contenant n boules numérotées de 1 à n (donc distinguables) et on effectue le tirage de r boules dans l'urne...

1. **sans remise et ordonnés (arrangements,** $r \leq n$ **car sans remise) :** il y en a $A_{n,r} = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$. Il s'agit du nombre de r-uplets $(b_1,\ldots,b_r)$ constitués d'éléments $b_1,\ldots,b_r$ de $\{1,\ldots,n\}$ deux à deux distincts. Pour $r = n$ on trouve le nombre de manières de permuter $\{1,\ldots,n\}$, soit $A_{n,n} = n!$.
   Exemple : nombre de tiercés avec n chevaux au départ ($r = 3$).
2. **sans remise et non ordonnés (combinaisons,** $r \leq n$ **car sans remise) :** il y en a[2] $\binom{n}{r} = \frac{A_{n,r}}{r!} = \frac{n!}{r!(n-r)!}$. Il s'agit du nombre de sous-ensembles de cardinal r de $\{1,\ldots,n\}$, autrement dit du nombre d'injections de $\{1,\ldots,r\}$ dans $\{1,\ldots,n\}$.
   Exemple : nombre de binômes possibles dans une classe de n élèves ($r = 2$).
3. **avec remise et ordonnés (**n**-uplets) :** il y en a n^r. Il s'agit du nombre de r-uplets $(b_1,\ldots,b_r)$ constitués d'éléments $b_1,\ldots,b_r$ de $\{1,\ldots,n\}^r$, c'est-à-dire le nombre d'applications de $\{1,\ldots,r\}$ dans $\{1,\ldots,n\}$. Bien entendu, $n^r \geq A_{n,r} \geq \binom{n}{r}$.
   Exemple : nombre de mots de r lettres ($n = 26$).
4. **avec remise et non ordonnés :** il y en a $\binom{n+r-1}{n-1}$. Il s'agit également du nombre de manières de placer r boules indistinguables dans n urnes distinguables, ou encore le nombre d'applications $f : \{1,\ldots,n\} \to \{1,\ldots,r\}$ vérifiant $f(1) + \cdots + f(n) = r$. Pour obtenir la formule, on aligne les r boules et on place $n-1$ cloisons délimitant les n urnes. Il y a $r+1$ possibilités pour la première cloison, $r+2$ pour la seconde, ..., $r+n-1$ possibilités pour la $n-1$ ième cloison. L'ordre de placement des $n-1$ cloisons n'étant pas pertinent, on obtient $(r+1)\cdots(r+n-1)/(n-1)! = \binom{n+r-1}{n-1}$.
   Exemple : nombre de possibilités au jeu des chiffres et des lettres ($n = 26$ et $r = 9$).

Exemple 2.4 (Tirage avec remise)**.** *Si on dispose d'une urne contenant* n *boules numérotées de* 1 *à* n *(donc distinguables), alors on modélise le tirage de* r *boules avec remise par la probabilité uniforme sur l'univers* $\Omega = \{1,\ldots,n\}^r$*, dont le cardinal vaut* n^r*. En conséquence, lors d'un tirage avec remise de deux cartes dans un jeu de* 32 *cartes, la probabilité que les cartes soient de la même couleur*[3] *vaut* $(2 \times 16^2)/32^2 = 1/2$*, tandis que la probabilité d'obtenir* 2 *as vaut* $4^2/32^2 = 1/64$.

2. On utilise la notation $\binom{n}{r}$ plutôt que C_n^r. Formule du binôme : $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.
3. Par *couleur* on entend ici de manière non standard *rouge* ou *noir* et pas *pique, carreau, trèfle, cœur*.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Exemple 2.5 (Tirage sans remise). *Si on dispose d'une urne contenant n boules numérotées de 1 à n (donc distinguables), alors on modélise le tirage de r boules sans remise par la probabilité uniforme sur l'univers $\Omega = \{T \subset \{1, \ldots, n\} : \mathrm{card}(T) = r\}$, dont le cardinal vaut $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. En conséquence, lors d'un tirage sans remise de deux cartes dans un jeu de 32 cartes, la probabilité de tirer deux cartes de même couleur vaut $2\binom{16}{2}/\binom{32}{2} = 15/31$, tandis que la probabilité d'obtenir 2 as vaut $\binom{4}{2}/\binom{32}{2} = 3/(32 \times 31)$. Comme on peut s'y attendre, ces deux probabilités sont plus petites que celles avec remise.*

Exemple 2.6 (Jeu de pile ou face). *Qu'elle est la probabilité d'obtenir k fois pile en n parties de pile ou face avec une pièce de monnaie équilibrée. On prend $0 \leq k \leq n$. L'univers est $\Omega = \{(a_1, \ldots, a_n) \in \{0,1\}^n\}$ de cardinal 2^n, où 0 code face et 1 code pile. L'évènement d'intérêt est $A = \{(a_1, \ldots, a_n) : a_1 + \cdots + a_n = k\}$, de cardinal $\binom{n}{k}$. La pièce étant équilibrée, on choisit le modèle d'équiprobabilité, et donc $\mathbb{P}(A) = \mathrm{card}(A)/\mathrm{card}(\Omega) = \binom{n}{k}2^{-n}$. Nous verrons qu'il s'agit d'un cas particulier de la loi binomiale (taille n et paramètre $1/2$).*

Exemple 2.7 (Erreur de modélisation). *Considérons le résultat du jet simultané de deux dés indistinguables. La modélisation par la probabilité uniforme sur $\{\{i,j\} : 1 \leq i,j \leq 6\}$ n'est pas bonne (elle donne trop de poids à l'égalité des dés). Le bon modèle est la probabilité uniforme sur $\{(i,j) : 1 \leq i,j \leq 6\}$, qui force à distinguer les dés, et donc à reformuler les évènements où ils ne sont pas distingués. Par exemple : pour tous $1 \leq i,j \leq 6$, l'évènement $\{i,j\} = \{(i,j),(j,i)\}$ a probabilité $1/21$ dans le mauvais modèle, tandis que dans le bon modèle, il a probabilité $2/36 = 1/18$ si $i \neq j$ et $1/36$ si $i = j$.*

Exemple 2.8 (Échantillonnage). *Considérons une urne contenant $N = N_1 + N_2$ boules dont N_1 blanches et N_2 rouges. On effectue un tirage sans remise de $n \leq N$ boules dans l'urne. Il y a $\binom{N}{n}$ tirages possibles. Adoptons le modèle de la probabilité uniforme sur l'ensemble de ces possibilités, c'est-à-dire sur l'univers Ω des sous-ensembles de $\{1, \ldots, N\}$ à n éléments. Pour tout $0 \leq k \leq n$, le nombre de tirages avec k boules blanches est $\binom{N_1}{k}\binom{N_2}{n-k}$, et la probabilité de tirer k boules blanches vaut donc*

$$\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

Cette formule définit la loi hypergéométrique sur les sous-populations de taille n d'une population de taille N à deux types (exemple 3.8). Cela montre au passage que

$$\binom{N}{n} = \sum_{k=0}^{n} \binom{N_1}{k}\binom{N_2}{n-k}.$$

Il s'agit de l'identité de convolution de Vandermonde[4]. La formule de Stirling[5] entraîne que $\binom{N_1}{k}\binom{N_2}{n-k}/\binom{N}{n} \to p^k(1-p)^{n-k}\binom{n}{k}$ lorsque $N_1, N_2 \to \infty$ avec $N_1/N \to p$. Cette formule définit la loi binomiale de taille n et de paramètre p. La généralisation à $d \geq 2$ types se fait comme suit : l'urne contient $N = N_1 + \cdots + N_d$ boules dont N_i boules de couleur i pour tout $1 \leq i \leq d$. On effectue un tirage sans remise de n boules. Il y a $\binom{N}{n}$ tirages possibles. Pour tous $n_1, \ldots, n_d$ vérifiant $n_i \leq N_i$ pour tout $1 \leq i \leq d$ et $n_1 + \cdots + n_d = n$, le nombre de tirages comportant n_i boules de couleur i pour tout $1 \leq i \leq d$ est $\binom{N_1}{n_1} \cdots \binom{N_d}{n_d}$. Avec le modèle de la probabiltié uniforme sur les tirages, la probabilité du tirage $(n_1, \ldots, n_d)$ est

$$\frac{\binom{N_1}{n_1} \cdots \binom{N_d}{n_d}}{\binom{N}{n}}.$$

4. Découle également de $(1+x)^{N_1}(1+x)^{N_2} = (1+x)^N$ en développant et en identifiant.
5. $n! \sim \sqrt{2\pi n}(n/e)^n$ ou plus précisément $n! = \sqrt{2\pi n}(n/e)^n(1 + \mathcal{O}_{n\to\infty}(1/n))$.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Cette formule définit la loi hypergéométrique « multitype » (exemple 4.3). On retrouve au passage une version multivariée de l'identité de convolution Vandermonde[6] *:*

$$\binom{N}{n} = \sum_{\substack{n_1+\cdots+n_d=n \\ n_1\le N_1,\dots,n_d\le N_d}} \binom{N_1}{n_1}\cdots\binom{N_d}{n_d}.$$

La formule de Stirling indique que $\binom{N_1}{n_1}\cdots\binom{N_d}{n_d}/\binom{N}{n}$ *converge vers* $\frac{n!}{n_1!\cdots n_d!}p_1^{n_1}\cdots p_d^{n_d}$ *lorsque* $N_1,\dots,N_d\to\infty$ *avec* $(N_1/N,\dots,N_d/N)\to(p_1,\dots,p_d)$*. Cette formule définit la loi multinomiale de taille* n *et de paramètre* $(p_1,\dots,p_d)$ *(exemple 4.2).*

Remarque 2.9 (Tirage aléatoire d'une partie). *Le tirage d'une partie à* k *élément dans un ensemble à* n *éléments est modélisé par la loi uniforme sur les* $\binom{n}{k}$ *parties à* k *éléments. Il est également possible d'effectuer ce tirage élément par élément, ce qui correspond à* k *tirages sans remise successifs :*

$$k!\frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-k+1} = \frac{1}{\binom{n}{k}}.$$

Le membre de droite provient du fait qu'il y a $k!$ *manières d'ordonner les* k *éléments à tirer, et la probabilité de chaque suite de* k *tirages vaut* $1/(n(n-1)\cdots(n-k+1))$*.*

Exemple 2.10 (Échantillon). *On tire* 4 *cartes dans un jeu de* 32*. Qu'elle est la probabilité d'obtenir* 2 *as et* 2 *rois ? Avec le modèle de l'exemple 2.8 avec* $N=32$*,* $(N_1,N_2,N_3)=(4,4,24)$ *(as, rois, autres cartes),* $(n_1,n_2,n_3)=(2,2,0)$*. La probabilité vaut*

$$\frac{\binom{2}{4}\binom{2}{4}\binom{0}{24}}{\binom{32}{4}} \approx 0,001 \quad \textit{(une chance sur mille).}$$

Autre exemple : dans un village de N *habitants adultes, dont* N_1 *retraités,* N_2 *actifs, et* N_3 *chômeurs, on tire au hasard les* n *membres du comité des fêtes. Qu'elle est la probabilité d'obtenir* n_1 *retraités,* n_2 *actifs, et* n_3 *chômeurs ? (*$n\le N$*,* $n_1\le N_1$*,* $n_2\le N_2$*,* $n_3\le N_3$*).*

Exemple 2.11 (Problème des anniversaires). *Calculons la probabilité* p_n *pour que dans une classe de* n *élèves, au moins deux d'entre eux soient nés le même jour. On suppose pour simplifier que les naissances sont uniformes sur les jours de l'année, et on ne tient pas compte des années bissextiles. On modélise cette expérience par la probabilité uniforme sur l'univers* $\Omega=\{1,\dots,d\}^n$ *où* $d=365$*, dont le cardinal est* d^n*. Cela correspond à faire* n *tirages avec remise dans une urne contenant* d *boules numérotés de* 1 *à* d*. Si* A *est l'évènement « deux élèves au moins sont nés le même jour » alors* A^c *correspond à* n *tirages sans remise ordonnés (arrangements !) et donc (pour* $n\le d$ *car* $p_n=1$ *sinon)*

$$p_n = \mathbb{P}(A) = 1-\mathbb{P}(A^c) = 1-\frac{d(d-1)\cdots(d-n+1)}{d^n} = 1-\prod_{k=1}^{n-1}\left(1-\frac{k}{d}\right).$$

La suite $(p_n)_{n\ge1}$ *croît de manière sigmoïde, et on a un phénomène de seuil (mode de la pente) vers* $20-30$ *car* $p_{20}\approx0,41$*,* $p_{30}\approx0,71$*,* $p_{40}\approx0,89$*,* $p_{50}\approx0,97$*. Ceci s'explique par le fait que le nombre de couple d'élèves est quadratique en* n*. Par contraste, si* B *est l'évènement « un élève au moins est né le même jour que l'enseignant » alors on a*

$$\mathbb{P}(B) = 1-\mathbb{P}(B^c) = 1-\left(\frac{d-1}{d}\right)^n = 1-\left(1-\frac{1}{d}\right)^n,$$

formule qui ne fait par apparaître de phénomène de seuil. Cette fois-ci, le nombre de couples (élève,enseignant) est linéaire en n*.*

6. Découle également de $(1+x)^{N_1}\cdots(1+x)^{N_d}=(1+x)^N$ en développant et en identifiant !

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Exemple 2.12 (Problème du chevalier de Méré). *Si l'on jette* 4 *fois un dé à six faces, la probabilité d'obtenir un* 6 *vaut* $1-(5/6)^4 \approx 0,52 > 1/2$. *Si l'on jette* 24 *fois deux dés à six faces, la probabilité d'obtenir un double six vaut* $1-(35/36)^{24} \approx 0,49 < 1/2$. *Le chevalier de Méré était un noble de la cour de Louis XIV, qui trouvait ces résultats contre intuitifs car il formalisait mal le problème (Pascal en fait part à Fermat dans l'une de ses lettres). Dans le premier cas, on utilise ici l'équiprobabilité sur l'univers* $\{1,\dots,6\}^4$ *et dans le second cas, l'équiprobabilité sur l'univers* $(\{1,\dots,6\}^2)^{24} = \{(i,j) : 1 \leq i,j \leq 6\}^{24}$.

Exemple 2.13 (Loi uniforme sur le groupe symétrique). *L'ensemble* $\mathcal{S}_n$ *des permutations de* $\{1,\dots,n\}$ *muni de la composition* $\circ$ *constitue ce qu'on appelle le* groupe symétrique. *Il s'agit d'un groupe fini non abélien de cardinal* $n!$. *La loi uniforme* μ *sur* $\mathcal{S}_n$ *qui affecte la probabilité* $1/n!$ *à chaque atome de* $\mathcal{S}_n$ *est la seule loi sur* $\mathcal{S}_n$ *invariante par toute translation (à droite, ou à gauche). La condition est évidemment nécessaire, et sa suffisance s'établit en observant que si* μ *est invariante par toute translation (disons à gauche) alors* $\mu(\sigma \circ \sigma') = \mu(\sigma')$ *pour tous* $\sigma, \sigma' \in \mathcal{S}_n$ *et donc* μ *affecte la même probabilité à tous les atomes de* $\mathcal{S}_n$ *(en posant* $\sigma^{-1} = \sigma'$*) ce qui entraîne que* μ *est la loi uniforme sur* $\mathcal{S}_n$.

Exemple 2.14 (Probabilité produit). *Si* $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ *et* $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ *sont deux espaces probabilisés, on admet l'existence et l'unicité d'une mesure de probabilité notée* $\mathbb{P}_1 \otimes \mathbb{P}_2$ *sur le produit* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, *appelée* **mesure de probabilité produit**, *qui vérifie*

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$$

pour tout évènement produit $A_1 \times A_2 \in \mathcal{F}_1 \times \mathcal{F}_2$. *De même, si* $(\Omega, \mathcal{F}, \mathbb{P})$ *est un espace de probabilité, on équipe* $\Omega^{\mathbb{N}}$ *de la tribu des cylindres* $\mathcal{F}^{\otimes\infty}$ *et on admet l'existence et l'unicité d'une unique mesure de probabilité* $\mathbb{P}^{\otimes\infty}$ *sur* $(\Omega^{\mathbb{N}}, \mathcal{F}^{\otimes\infty})$ *qui vérifie*

$$\mathbb{P}^{\otimes\infty}(A_0 \times A_1 \times A_2 \times \cdots) = \mathbb{P}(A_0)\mathbb{P}(A_1)\mathbb{P}(A_2)\cdots$$

pour tout cylindre $A_0 \times A_1 \times A_2 \cdots \in \mathcal{F}^{\otimes\infty}$ *(il s'agit à droite d'un produit fini car* $\mathbb{P}(A_n) = 1$ *à partir d'un certain rang sur* n*). Cette construction est utilisée pour le jeu de pile ou face.*

Remarque 2.15 (Produit et équiprobabilité). *Si* $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ *et* $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ *vérifient l'équiprobabilité alors* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ *vérifie aussi l'équiprobabilité.*

Théorème 2.16 (Suites). *Pour toute suite* (A_n) *d'évènements sur* $(\Omega, \mathcal{F}, \mathbb{P})$,

1. *si* (A_n) *est croissance (pour l'inclusion) alors* $\mathbb{P}(A_n) \nearrow \mathbb{P}(\cup_n A_n)$
2. *si* (A_n) *est décroissance (pour l'inclusion) alors* $\mathbb{P}(A_n) \searrow \mathbb{P}(\cap_n A_n)$
3. $\mathbb{P}(\cup_n A_n) \leq \sum_n \mathbb{P}(A_n)$
4. *si* $\mathbb{P}(A_n) = 0$ *pour tout* n *alors* $\mathbb{P}(\cup_n A_n) = 0$
5. *si* $\mathbb{P}(A_n) = 1$ *pour tout* n *alors* $\mathbb{P}(\cap_n A_n) = 1$

Démonstration. Pour la première propriété, les $B_n = A_n \setminus A_{n-1}$ sont deux à deux disjoints, d'où $\mathbb{P}(\cup_n A_n) = \mathbb{P}(\cup_n B_n) = \sum_n \mathbb{P}(B_n) = \lim_n \sum_{m \leq n} \mathbb{P}(B_m) = \lim_n \mathbb{P}(A_n)$. □

Il est commode d'interpréter $\mathbb{P}(A)$ comme la surface (ou le cardinal) du patatoïde A dessiné sur le plan (ou sur le réseau $\mathbb{Z}^2$). Le théorème suivant, connu également sous le nom de *crible de Poincaré*, nous dit que la surface d'une union est égale à la somme des surfaces, moins la surface des intersections deux à deux, plus la surface des intersections trois à trois, etc, ce qui correspond à inclure et exclure alternativement.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Théorème 2.17 (Principle d'inclusion-exclusion). *Pour tous* $A_1, \ldots, A_r \in \mathcal{F}$

$$\mathbb{P}(\cup_{1\le i\le r} A_i) = \sum_{k=1}^{r} (-1)^{k+1} S_k \quad \text{où} \quad S_k = \sum_{1\le i_1<\cdots<i_k\le r} \mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_r}).$$

Pour $r = 2$, on retrouve $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$, et pour $r = 3$,

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)\\ - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Le principe d'inclusion-exclusion est rarement utilisé pour $r > 2$. Il l'est cependant dans l'étude de la fluctuation asymptotique du collectionneur de coupons (théorème A.14).

Démonstration. On procède par récurrence sur r, en observant que

$$\begin{aligned}\mathbb{P}(\cup_{1\le i\le r+1} A_i) &= \mathbb{P}(\cup_{1\le i\le r} A_i) + \mathbb{P}(A_{r+1}) - \mathbb{P}((\cup_{1\le i\le r} A_i) \cap A_{r+1})\\ &= \mathbb{P}(\cup_{1\le i\le r} A_i) + \mathbb{P}(A_{r+1}) - \mathbb{P}(\cup_{1\le i\le r}(A_i \cap A_{r+1}))\end{aligned}$$

ce qui permet d'utiliser l'hypothèse de récurrence (pour le premier et dernier terme). □

2.1 Conditionnement

Intuitivement, une mesure de probabilité permet de quantifier le manque d'information en affectant un poids aux issues possibles. La prise en compte d'une information supplémentaire s'effectue en restreignant la mesure de probabilité, c'est le conditionnement.

Définition 2.18 (Probabilité conditionnelle). *Si* $A, B \in \mathcal{F}$ *avec* $\mathbb{P}(B) > 0$ *alors la* probabilité conditionnelle de A sachant B *est la quantité (parfois notée* $\mathbb{P}_B(A)$*) suivante :*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Théorème 2.19 (Propriétés importantes).

1. *si* $\mathbb{P}(B) > 0$ *alors l'application*

$$\mathbb{P}(\cdot|B) : A \in \mathcal{F} \to \mathbb{P}(A|B) \in [0, 1]$$

est une mesure de probabilité sur $(\Omega, \mathcal{F})$ *appelée* probabilité conditionnelle. *En particulier* $\mathbb{P}(\Omega|B) = 1$, *et pour toute suite* (A_n) *d'évènements disjoints,*

$$\mathbb{P}(\cup_n A_n|B) = \sum_n \mathbb{P}(A_n|B)$$

2. *si* $0 < \mathbb{P}(B) < 1$ *alors*

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

et plus généralement si $\Omega = \cup_n B_n$ *est une partition de* Ω *avec* $\mathbb{P}(B_n) > 0$ *alors*

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

On a $\mathbb{P}(B|B) = 1$ et $\mathbb{P}(A|B) = 0$ si $A \cap B = \varnothing$ de sorte que $\mathbb{P}(\cdot|B)$ est portée par B en quelque sorte. En fait, $\mathbb{P}(\cdot|B)$ est une mesure de probabilité sur $(B, \mathcal{F}_B)$ où $\mathcal{F}_B = \{C \cap B : B \in \mathcal{F}\}$ est la tribu trace de $\mathcal{F}$ sur B. Notons que $\mathbb{P}(\cdot|\Omega) = \mathbb{P}$.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Remarque 2.20 (Erreur fréquente). $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$ *mais* $\mathbb{P}(A|B^c) \neq 1 - \mathbb{P}(A|B)$.

La formule de Bayes permet d'inverser l'ordre d'un conditionnement.

Théorème 2.21 (Formule de Bayes). *Si* $\mathbb{P}(A) > 0$ *et* $0 < \mathbb{P}(B) < 1$ *alors*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

La formule de Bayes permet de transformer les données $\mathbb{P}(A|B), \mathbb{P}(A|B^c), \mathbb{P}(B)$ en $\mathbb{P}(B|A)$. Notons la formule suivante parfois utile en pratique pour les calculs numériques :

$$\mathbb{P}(B|A) = \frac{1}{1 + \frac{\mathbb{P}(A|B^c)}{\mathbb{P}(A|B)} \frac{\mathbb{P}(B^c)}{\mathbb{P}(B)}}.$$

Exemple 2.22 (Test de dépistage de maladie). *Des laboratoires pharmaceutiques ont mis au point un test médical pour dépister une maladie. Les experts pensent qu'une personne sur mille est malade dans la population. De plus, des expériences ont montré que le test déclare positifs* 99% *des malades qu'on lui soumet, et qu'il déclare malades* 2% *des personnes saines qu'on lui soumet. Si on définit les évènements*

$$A = \text{« le test médical est positif »} \quad \textit{et} \quad B = \text{« la personne est malade »}$$

alors les données se traduisent par $\mathbb{P}(B) = 1/1000$, $\mathbb{P}(A|B) = 99/100$, *et* $\mathbb{P}(A|B^c) = 2/100$. *La probabilité que la personne soit malade sachant que test est positif vaut*

$$\mathbb{P}(B|A) = \frac{1}{1 + \frac{\mathbb{P}(A|B^c)}{\mathbb{P}(A|B)} \frac{\mathbb{P}(B^c)}{\mathbb{P}(B)}} = \frac{1}{1 + \frac{2 \times 999}{99}} \approx \frac{1}{20} = 0,05.$$

Le test n'est vraiment pas efficace de ce point de vue ! Le paradoxe vient du fait que sur 99.9% *de la population le test se trompe dans* 2% *des cas, et que ces cas représentent finalement une grande part des réponses positives du test. Cet exemple est l'occasion de rappeler que du point de vue statistique, un test comporte deux types d'erreur (faux positifs et faux négatifs) qui ne jouent pas un rôle symétrique du point de vue du risque modélisé.*

2.2 Indépendance

Intuitivement, deux expériences sont indépendantes lorsqu'elles ne sont pas reliées par une relation causale. Par exemple, le jet de deux dés équilibrés, de couleurs différentes, à six faces, est modélisés par l'univers $\Omega \times \Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. L'absence de relation causale entre les deux dés suggère de considérer la mesure de probabilité uniforme sur $(\Omega \times \Omega, \mathcal{P}(\Omega \times \Omega))$, qui se trouve être la mesure de probabilité produit des mesures de probabilités uniformes. En particulier, si $A \times \Omega$ (respectivement $\Omega \times B$) est un évènement qui ne concerne que le résultat du premier (respectivement second) jet de dé, alors[7]

$$\begin{aligned}
\mathbb{P}((A \times \Omega) \cap (\Omega \times B)) &= \frac{\mathrm{Card}((A \times \Omega) \cap (\Omega \times B))}{\mathrm{Card}(\Omega \times \Omega)} \\
&= \frac{\mathrm{Card}(A)\mathrm{Card}(B)}{\mathrm{Card}(\Omega)\mathrm{Card}(\Omega)} \\
&= \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A \times \Omega)\mathbb{P}(\Omega \times B)
\end{aligned}$$

7. Sur le plan cartésien $\Omega \times \Omega$, l'évènement $A \times \Omega$ est une bande verticale et $\Omega \times B$ une bande horizontale.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

(notons au passage l'abus qui consiste à noter $\mathbb{P}$ à la fois la mesure de probabilité uniforme sur Ω et sur $\Omega \times \Omega$). Cette formule remarquable conduit à la définition générale suivante de l'indépendance, bien au delà du cas de l'équiprobabilité.

Définition 2.23 (Indépendance de deux évènements). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On dit que les* ***deux évènements*** *$A, B \in \mathcal{F}$ sont* ***indépendants*** *lorsque*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Exemple 2.24. *Considérons le lancer d'un dé équilibré, modélisé par $\Omega = \{1, 2, 3, 4, 5, 6\}$ équipé de la tribu de toutes les parties $\mathcal{P}(\Omega)$ et de la mesure de probabilité uniforme. Soit $A =$«le résultat est ≤ 4», $B =$«le résultat est pair» alors A et B sont indépendants car en utilisant les cardinaux : $\mathbb{P}(A \cap B) = \frac{2}{6} = \frac{4}{6}\frac{3}{6} = \mathbb{P}(A)\mathbb{P}(B)$.*

Remarque 2.25. *Si A et B sont indépendants alors A^c et B^c le sont car*

$$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A \cup B) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)).$$

En revanche, si A et B sont indépendants et si $A' \subset A$ alors A' et B ne sont pas forcément indépendants. En effet, si par exemple $A' \subset B$ avec A' et B indépendants alors on aurait

$$\mathbb{P}(A') = \mathbb{P}(A' \cap B) = \mathbb{P}(A')\mathbb{P}(B)$$

ce qui est impossible si $\mathbb{P}(A') \neq 0$ et $\mathbb{P}(B) \neq 1$. Exemple : jet de dé équilibré à six faces, $A =$«le résultat est ≤ 4», $A' =$«le résultat vaut 2», et $B =$«le résultat est pair».

Définition 2.26 (Indépendance d'une famille d'évènements). *Si $(A_i)_{i \in I}$ est une* ***famille d'évènements****, on dit qu'ils sont* ***indépendants***[8] *lorsque pour toute partie* ***finie*** *$J \subset I$,*

$$\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j).$$

De même, si $(\mathcal{F}_i)_{i \in I}$ est une ***famille de tribus*** *sur Ω, on dit qu'elles sont indépendantes lorsque $(A_i)_{i \in I}$ sont indépendants dès que $A_i \in \mathcal{F}_i$ pour tout $i \in I$.*

Remarque 2.27 (Lien avec indépendance deux à deux). *L'indépendance implique clairement l'indépendance deux à deux mais la réciproque est fausse en général. Considérons par exemple le jet de deux dés parfaits de couleurs différentes, à six faces, et les évènements*

1. *$A_1 =$«le premier dé est pair»*
2. *$A_2 =$«le second dé est pair»*
3. *$A_3 =$«la somme des deux dés est impaire».*

On a alors en utilisant le modèle d'équiprobabilité sur $\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{2}\frac{1}{2}\frac{1}{2} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

tandis que $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = \frac{1}{4}$. Un autre contre exemple est donné par le jeu de pile ou face avec une pièce équilibrée (deux lancers suffisent) modélisé par l'équiprobabilité sur l'univers $\Omega = \{0, 1\} \times \{0, 1\} = \{(0, 0), (1, 1), (0, 1), (1, 0)\}$ de cardinal 4 avec 0 codant face et 1 codant pile. On considère alors les trois évènements suivants :

- *$A =$«le premier lancer donne face» $= \{(0, 0), (0, 1)\}$*
- *$B =$«le second lancer donne pile» $= \{(1, 1), (0, 1)\}$*

8. On dit aussi (rarement dans la pratique mais souvent dans la définition) *mutuellement indépendants*.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

– $C =$ «*les deux premiers lancers donne le même résultat*» $= \{(0,0),(1,1)\}$.
On a alors $A \cap B = \{(0,1)\}$, $A \cap C = \{(0,0)\}$, $B \cap C = \{(1,1)\}$, $A \cap B \cap C = \varnothing$. *Donc*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(C) = \mathbb{P}(B)\mathbb{P}(C)$$

tandis que $\mathbb{P}(A \cap B \cap C) = 0 \neq \frac{1}{2}\frac{1}{2}\frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. *Les évènements* A, B, C *sont deux à deux indépendantes mais ne sont pas (mutuellement) indépendants.*

Exemple 2.28 (Non transitif !). *Si* A *et* B *sont indépendants et* B *et* C *sont indépendants, alors* A *et* C *ne sont pas forcément indépendants ! Considérons par exemple le jet d'un dé équilibré à* 8 *faces, modélisé par le modèle équiprobable sur* $\Omega = \{1,2,3,4,5,6,7,8\}$. *Soit*

$$A = \{1,2,7,8\}, \quad B = \{2,3,6,8\}, \quad C = \{3,4,5,8\}.$$

La formule des cardinaux donne $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ *et*

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \frac{1}{4} \quad \text{et} \quad \mathbb{P}(A \cap B \cap C) = \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C),$$

mais $\mathbb{P}(A \cap C) = \frac{1}{8} \neq \mathbb{P}(A)\mathbb{P}(C)$. *L'indépendance est symétrique mais pas transitive.*

Théorème 2.29 (Propriétés importantes).

1. *si les* $(A_i)_{i \in I}$ *sont indépendants alors les* $(A_i^c)_{i \in I}$ *le sont aussi*
2. A *est indépendant de lui même si et seulement si* $\mathbb{P}(A) \in \{0,1\}$

Démonstration. Pour la première propriété, on raisonne par récurrence à partir de

$$\begin{aligned}
\mathbb{P}(A^c \cap B^c) &= 1 - \mathbb{P}(A \cup B) \\
&= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) \\
&= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) \\
&= (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A^c)\mathbb{P}(B^c).
\end{aligned}$$

□

Remarque 2.30 (Erreur fréquente). *Ne pas confondre* A *et* B ***incompatibles*** *avec* A *et* B ***indépendants***. *La première notion est purement ensembliste, tandis que la seconde nécessite une mesure de probabilité. Notons que si* A *et* B *sont à la fois indépendants et incompatibles alors* $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(\varnothing) = 0$ *et donc* $\mathbb{P}(A) = 0$ *ou* $\mathbb{P}(B) = 0$. *Notons enfin que si* A *et* B *sont indépendants alors* A^c *et* B^c *le sont, tandis que si* A *et* B *sont incompatibles, alors* A^c *et* B^c *ne le sont que s'ils forment une partition de* Ω

Lemme 2.31 (Borel-Cantelli). *Soit* (A_n) *une suite d'évènements dans* $(\Omega, \mathcal{F}, \mathbb{P})$.

1. *Cantelli : si* $\sum_n \mathbb{P}(A_n) < \infty$ *alors* $\mathbb{P}(\overline{\lim}\, A_n) = 0$
2. *Borel (loi du zéro-un) : si les* (A_n) *sont* ***indépendants*** *alors*

$$\mathbb{P}(\overline{\lim}\, A_n) = \begin{cases} 0 & \text{ssi } \sum_n \mathbb{P}(A_n) < \infty \\ 1 & \text{ssi } \sum_n \mathbb{P}(A_n) = \infty. \end{cases}$$

La seconde partie du lemme contient une réciproque à la première partie.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. Pour la partie Cantelli : la suite (B_n) définie par $B_n = \cup_{m\geq n} A_m$ est décroissante. Par conséquent, si $\sum_n \mathbb{P}(A_n) < \infty$ alors

$$\mathbb{P}(\overline{\lim}\, A_n) = \mathbb{P}(\cap_n B_n) = \lim_n \mathbb{P}(B_n) \leq \lim_n \sum_{m\geq n} \mathbb{P}(A_m) = 0.$$

Pour la partie Borel : la première partie réduit le problème à établir que si $\sum_n \mathbb{P}(A_n) = \infty$ alors $\mathbb{P}(\overline{\lim}\, A_n) = 1$, ou encore $\mathbb{P}(\underline{\lim}\, A_n^c) = 0$. Si $B_n = \cap_{m\geq n} A_m^c$ alors (B_n) est croissante et donc $\mathbb{P}(\underline{\lim}\, A_n^c) = \lim_n \mathbb{P}(B_n)$. À présent, l'indépendance des (A_n^c), l'inégalité $1 - x \leq e^{-x}$ valable pour $x \in \mathbb{R}$, et $\sum_n \mathbb{P}(A_n) = \infty$, donnent pour tout n :

$$\mathbb{P}(B_n) = \prod_{m\geq n} \mathbb{P}(A_n^c) = \prod_{m\geq n} e^{-\mathbb{P}(A_m)} = e^{-\sum_{m\geq n} \mathbb{P}(A_m)} = 0.$$

□

Remarque 2.32 (À quoi peut bien servir le lemme de Borel-Cantelli ?). *La première partie permet par exemple d'établir une version de la loi forte des grands nombres (Théorème 5.4), tandis que la seconde partie permet par exemple d'établir que tout mot fini se répète une infinité de fois dans un jeu de pile ou face non dégénéré (Remarque A.2).*

Si $\mathcal{F}, \mathcal{F}'$ sont des tribus sur Ω avec $\mathcal{F}' \subset \mathcal{F}$ alors $\mathcal{F}'$ est une *sous-tribu* de $\mathcal{F}$.

Théorème 2.33 (Loi du zéro-un de Kolmogorov[9]). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(\mathcal{F}_n)_{n\geq 1}$ une suite de sous-tribus de $\mathcal{F}$, **indépendantes**. Soit $\mathcal{G}_n$ la tribu engendrée par $\cup_{m\geq n}\mathcal{F}_m$. Alors $\mathbb{P}(A) \in \{0, 1\}$ pour tout A dans la* tribu terminale $\mathcal{G}_\infty = \cap_n \mathcal{G}_n$.

En quelque sorte, on a $\mathcal{G}_\infty = \overline{\lim}\, \mathcal{F}_n$. Notons que si $A_n \in \mathcal{F}_n$ pour tout n alors $\underline{\lim}_n A_n$ et $\overline{\lim}_n A_n$ sont dans la tribu terminale et donc sont de probabilité 0 ou 1, ce qui montre que le second lemme de Borel-Cantelli découle de la loi du zéro-un de Kolmogorov.

Démonstration. Les tribus $\mathcal{F}_n$ et $\mathcal{G}_{n+1}$ sont indépendantes, et donc la tribu engendrée par $\cup_n \mathcal{F}_n$ est indépendante de la tribu $\mathcal{G}_\infty$. Comme $\mathcal{G}_\infty \subset \cup_n \mathcal{F}_n$ on en déduit que tout évènement de $\mathcal{G}_\infty$ est indépendant de lui même, et est donc de probabilité 0 ou 1. □

9. Hors programme.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Chapitre 3

Variables aléatoires réelles

Définition 3.1 (Variable aléatoire réelle). *Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé, on appelle* variable aléatoire réelle *(abrégé v.a.r.) toute application $X : \Omega \to \mathbb{R}$ telle que*

$$\{X \in I\} = \{\omega \in \Omega : X(\omega) \in I\} = X^{-1}(I) \in \mathcal{F}$$

pour tout intervalle $I \subset \mathbb{R}$. On dit que X est une v.a.r. discrète *lorsque $X(\Omega)$ est fini ou infini dénombrable, typiquement $X(\Omega) = \mathbb{N}$.*

Exemple 3.2. *Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité et $A \in \mathcal{F}$ alors $\mathbf{1}_A$ est une variable aléatoire discrète booléenne (i.e. prenant les valeurs 0 ou 1). On dit qu'il s'agit d'une variable aléatoire de Bernoulli de paramètre $\mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$. Plus généralement, si $x_1, \ldots, x_n \in \mathbb{R}$ et $A_1, \ldots, A_n \in \mathcal{F}$ alors $x_1 \mathbf{1}_{A_1} + \cdots + x_n \mathbf{1}_{A_n}$ est une v.a.r. discrète.*

On rappelle que $\mathbb{R}$ est équipé de la tribu borélienne, engendrée par les intervalles. Une fonction $f : \mathbb{R} \to \mathbb{R}$ est borélienne lorsque $f^{-1}(I)$ appartient à la tribu borélienne de $\mathbb{R}$ pour tout intervalle $I \subset \mathbb{R}$. On admet que toute fonction continue est borélienne. On peut montrer que si X est une v.a.r. et si f est borélienne alors $f(X)$ est une v.a.r. On admet que si X et Y sont deux v.a.r. définies sur le même espace de probabilité alors la somme $X + Y$ et le produit XY sont des v.a.r. Par commodité, on note

$$\mathbb{P}(X \in I, Y \in J) = \mathbb{P}(\{X \in I\} \cap \{Y \in J\}).$$

Proposition 3.3 (Caractérisation). *Une application $X : (\Omega, \mathcal{F}) \to \mathbb{R}$ est une v.a.r. si et seulement si $\{X \leq x\} = X^{-1}(]-\infty, x]) \in \mathcal{F}$ pour tout $x \in \mathbb{R}$.*

Démonstration. Découle des axiomes des tribus car tout intervalle de $\mathbb{R}$ s'obtient en utilisant un nombre au plus dénombrable d'intervalles de la forme $]-\infty, x]$ et symboles $\cap, \cup, ^c$. Par exemple $]a, b] =]-\infty, b] \cap]-\infty, a]^c$ et $[a, b] = \cap_{n \in \mathbb{N}^*}]a - 1/n, b]$. □

3.1 Fonction de répartition et loi

Définition 3.4 (Fonction de répartition). *La fonction de répartition F_X d'une v.a.r. X est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie par $F_X(x) = \mathbb{P}(X \leq x)$ pour tout $x \in \mathbb{R}$.*

Théorème 3.5 (Propriété des fonction de répartition). *Si X est une v.a.r. alors*

1. *F_X est croissante et continue à droite*
2. *$\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$*
3. *$F_X(x^-) = \mathbb{P}(X < x)$ pour tout $x \in \mathbb{R}$*

Copyrighted material

Démonstration. La fonction F_X est croissante car si $x \leq y$ alors $\{X \leq x\} \subset \{X \leq y\}$. Les propriétés suivantes s'obtiennent en utilisant le théorème 2.16 avec...

- l'union croissante $\cup_n \{X \leq x_n\} = \mathbb{R}$ pour toute suite $(x_n) \nearrow \infty$
- l'intersection décroissante $\cap_n \{X \leq x_n\} = \varnothing$ pour toute suite $(x_n) \searrow -\infty$
- l'intersection décroissante $\cap_n \{X \leq x_n\} = \{X \leq x\}$ pour toute suite $(x_n) \searrow x$
- l'union croissante $\{X < x\} = \cup_n \{X \leq x_n\}$ pour toute suite $(x_n) \nearrow x$ avec $x_n < x$

□

Définition 3.6 (Médiane, quartiles, quantiles). *Si X est une v.a.r. de fonction de répartition F_X, alors pour tout $\alpha \in [0,1]$, et tout $x \in \mathbb{R}$, on dit que x est un quantile d'ordre α si $F_X(x) = \alpha$. Pour $\alpha = \frac{1}{2}$ on parle de médiane. On appelle quartiles les quantiles d'odre $\frac{1}{4}, \frac{2}{4} = \frac{1}{2}, \frac{3}{4}$. La médiane en fait partie. La notion de quantile permet de décrire grossièrement la répartition des valeurs de la v.a.r. X : salaire médian, poids médian, âge médian, etc. Elle joue un rôle important dans la construction des tests statistiques.*

Définition 3.7 (Loi). *La loi d'une v.a.r. X est la mesure de probabilité $\mathbb{P}_X$ sur $\mathbb{R}$ définie pour tout intervalle $I \subset \mathbb{R}$ par $\mathbb{P}_X(I) = \mathbb{P}(X \in I)$. Si μ est une mesure de probabilité sur $\mathbb{R}$, on note $X \sim \mu$ et on dit que « X suit la loi μ » lorsque la v.a.r. X a pour loi μ. Le terme « loi » est synonyme de « mesure de probabilité ».*

Exemple 3.8 (Lois discrètes). *La loi d'une v.a.r. discrète à valeur dans un ensemble au plus dénombrable E est caractérisée par la donnée de $\mathbb{P}(X = x)$ pour tout $x \in E$, car pour tout intervalle $I \subset \mathbb{R}$ on a, en raison du fait que $I \cap E$ est au plus dénombrable,*

$$\mathbb{P}(X \in I) = \mathbb{P}(X \in I \cap E) = \sum_{x \in I \cap E} \mathbb{P}(X = x).$$

Voici quelques lois discrètes classiques qui apparaissent dans le jeu de pile ou face :

1. *loi de Bernoulli : $E = \{0,1\}$ et $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p \in [0,1]$*
2. *loi binomiale : $E = \{0,1,\ldots,n\}$ et $\mathbb{P}(X = k) = \binom{k}{n} p^k (1-p)^{n-k}$ pour tout $k \in E$*
3. *loi de Poisson : $E = \mathbb{N}$ et $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ avec $\lambda \geq 0$, pour tout $k \in E$*
4. *loi géométrique : $E = \mathbb{N}^*$ et $\mathbb{P}(X = k) = (1-p)^{k-1} p$ avec $p \in [0,1]$, pour tout $k \in E$*

Loi hypergéométrique : si dans une population de $N = N_1 + N_2$ individus dont N_1 sont de type 1 et N_2 de type 2, on effectue un sondage sans remise sur $n \leq N$ individus, alors le nombre X d'individus de type 1, parmi les n individus tirés, suit la loi hypergéométrique $\mathrm{HyperGeom}(N_1, N_2, n)$ sur $E = \{0,1,\ldots,n\}$ donnée pour tout $0 \leq k \leq n$ par

$$\mathbb{P}(X = k) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

On retrouve la loi binomiale $\mathrm{Binom}(n,p)$ lorsque $N_1, N_2 \to \infty$ avec $N_1/N \to p$ (en utilisant la formule de Stirling). Nous avons déjà abordé ces aspects sous l'angle combinatoire dans l'exemple 2.8. Les probabilités de la loi géométrique sont parfois difficiles à évaluer à cause du comportement explosif de la factorielle, et dans la pratique, il est parfois bien commode d'utiliser l'approximation binomiale, ce qui revient à utiliser la formule de Stirling.

Exemple 3.9 (Lois uniformes). *On dit qu'une v.a.r. X suit la loi uniforme sur l'ensemble fini $\{1,2,\ldots,n\}$ lorsque $\mathbb{P}(X = k) = 1/n$ pour tout $1 \leq k \leq n$. Dans ce cas, F_X est constante sur les morceaux $]-\infty,0[, [0,1[, \ldots, [n-1,n[, [n,\infty[$ et y prend les valeurs $0, 1/n, \ldots, (n-1)/n$,*

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

1. *On dit qu'une v.a.r. X suit la loi uniforme sur $[0,1]$ lorsque $\mathbb{P}_X(I) = |I \cap [0,1]|$ pour tout intervalle $I \subset \mathbb{R}$. Dans ce cas, pour tout $x \in \mathbb{R}$,*

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in [0,1] \\ 1 & \text{si } x \geq 1. \end{cases}$$

Exemple 3.10 (Lois à densité). *On dit qu'une fonction continue par morceaux $f : \mathbb{R} \to \mathbb{R}$ est une densité (de probabilité) lorsqu'elle vérifie*

$$f \geq 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

On dit qu'une v.a.r. X possède une loi de densité f lorsque pour tout intervalle $I \subset \mathbb{R}$,

$$\mathbb{P}(X \in I) = \int_I f(x)\,dx.$$

Aussi F_X est la primitive de f valant 1 en $+\infty$. Voici quelques exemples de lois à densité :

1. *loi uniforme sur $[a,b]$: $f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$ avec $a < b$*
2. *loi exponentielle : $f(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$ avec $\lambda > 0$*
3. *loi normale (ou gaussienne) : $f(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{(x-m)^2}{2\sigma^2})$ avec $m \in \mathbb{R}$ et $\sigma > 0$*
4. *loi de Cauchy : $f(x) = \frac{1}{\pi(1+x^2)}$*
5. *loi de Pareto : $f(x) = \frac{1}{(1-a)(1+x)^a}\mathbf{1}_{\mathbb{R}_+}(x)$ avec $a > 1$*
6. *loi du χ^2, loi Gamma, loi Beta, etc.*

Un bon exercice consiste à calculer lorsque cela est possible la fonction de répartition.

Théorème 3.11 (Caractérisation de la loi par la fonction de répartition). *Si X et Y sont deux v.a.r. alors $\mathbb{P}_X = \mathbb{P}_Y$ si et seulement si $F_X = F_Y$.*

Démonstration. Reprendre la preuve de la proposition 3.3. □

Le résultat suivant permet de simuler une variable aléatoire à partir de sa fonction de répartition et d'un générateur de la loi uniforme (appelée communément rand).

Théorème 3.12 (Méthode d'inversion). *Si X est une v.a.r. alors $F_X^{-1}(U)$ est une v.a.r. de même loi que X, où U est une v.a.r. de loi uniforme sur $[0,1]$ et où $F_X^{-1} :]0,1[\to \overline{\mathbb{R}}$ est la* fonction réciproque généralisée de F_X *définie pour tout $p \in [0,1]$ par*

$$F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}.$$

Démonstration. Comme $\lim_{x\to\infty} F_X(x) = 1$ et $\lim_{x\to-\infty} F_X(x) = 0$ il vient $F_X^{-1}(p) \in \mathbb{R}$ pour tout $0 < p < 1$. Si $F_X(x) < p$ alors $F_X^{-1}(p) > x$, et comme F_X est continue à droite,

$$F_X(F_X^{-1}(p)) \geq p.$$

Donc $F_X(x) \geq p$ ssi $F_X^{-1}(p) \leq x$ pour tous x et $0 < p < 1$. Comme $\mathbb{P}(0 < U < 1) = 1$,

$$\mathbb{P}(F_X^{-1}(U) \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x)$$

pour tout $x \in \mathbb{R}$ ce qui montre que $F_X^{-1}(U)$ et X ont même loi. Le fait que $F_X^{-1}(p)$ puisse prendre les valeurs $\pm\infty$ pour $p = 0$ et $p = 1$ n'a pas d'importance car

$$\mathbb{P}(U = 0) = \mathbb{P}(U = 1) = 0.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

D'autre part, le fait que $F_X^{-1}(U)$ soit une v.a.r. découle du fait que

$$\{F_X^{-1}(U) \leq x\} \cap \{0 < U < 1\} = \{U \leq F_X(x)\} \cap \{0 < U < 1\}.$$

□

Remarque 3.13 (Cas continu). *On sait que $F_X(F_X^{-1}(p)) \geq p$. Comme $F_X(x) < p$ pour $F_X^{-1}(p) > x$, il en découle que si F_X est continue en $x \in \mathbb{R}$ alors $F_X^{-1}(F_X(x)) \leq x$ et donc $F_X^{-1}(F_X(x)) = x$. En particulier, si F_X est continue et strictement croissante alors c'est une bijection de $\mathbb{R}$ sur $]0,1[$ et F_X^{-1} est sa fonction réciproque. On peut aussi établir que si F_X est continue alors $F_X(X)$ suit la loi uniforme sur $[0,1]$.*

Exemple 3.14 (Simulation d'une loi discrète finie). *Soit U une v.a.r. uniforme sur $[0,1]$. Pour tout $p \in [0,1]$, la v.a.r. $\mathbf{1}_{\{U \leq p\}}$ suit la loi de Bernoulli de paramètre p. Plus généralement, soit $p_1, \ldots, p_n \in [0,1]$ avec $p_1 + \cdots + p_n = 1$. Posons $a_0 = 0$ et $a_i = p_1 + \cdots + p_i$ pour tout $1 \leq i \leq n$. La v.a.r. X à valeurs dans $\{1, \ldots, n\}$ qui vaut par définition i sur l'évènement $\{U \in [a_{i-1}, a_i]\}$ vérifie forcément $\mathbb{P}(X = i) = p_i$ pour tout $1 \leq i \leq n$.*

Exemple 3.15 (Simulation de la loi uniforme sur $[a,b]$). *Si U suit la loi uniforme sur $[0,1]$ alors pour tout $a < b$, la v.a.r. $(b-a)U + a$ suit la loi uniforme sur $[a,b]$.*

Exemple 3.16 (Simulation de la loi exponentielle). *Si U suit la loi uniforme sur $[0,1]$ alors $-\ln(1-U)/\lambda$ suit la loi exponentielle de paramètre λ. Comme $1-U$ et U ont même loi, on peut utiliser alternativement $-\ln(U)/\lambda$. Sur un ordinateur, la fonction $\ln$ dilate la discrétisation de U près de 0 et la précision sera mauvaise dans cette zone. Cependant, cette zone est peu probable car elle correspond à la queue de distribution de la loi exponentielle.*

Exemple 3.17 (Simulation de la loi de Cauchy). *Si U suit la loi uniforme sur $[0,1]$ alors $\tan(\pi U)$ suit la loi de Cauchy. Sur un ordinateur, la fonction $\tan$ dilate la discrétisation de U près de 1 et la précision sera mauvaise dans cette zone qui est probable.*

3.2 Espérance

Nous nous basons sur une définition axiomatique de l'espérance des v.a.r. positives.

Théorème 3.18 (Espérance des variables positives – Admis). *Soit $L_+(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $[0,\infty]$. Il existe une unique application $\mathbb{E} : L_+(\Omega, \mathcal{F}, \mathbb{P}) \mapsto [0,\infty]$ avec les propriétés suivantes (convention $0 \times \infty = 0$) :*

1. *$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$ et en particulier $\mathbb{E}(\mathbf{1}_\Omega) = 1$*
2. *$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ pour tous $X, Y \in L_+(\Omega, \mathcal{F}, \mathbb{P})$ et $a, b \in \mathbb{R}_+$*
3. *$\mathbb{E}(\lim_{n\to\infty} X_n) = \lim_{n\to\infty} \mathbb{E}(X_n)$ pour toute suite (X_n) croissante de $L_+(\Omega, \mathcal{F}, \mathbb{P})$*

La troisième propriété est connue sous le nom de *convergence monotone*.

Démonstration. Si $X = \sum_{k=1}^n x_k \mathbf{1}_{A_k}$ avec $A_1, \ldots, A_n \in \mathcal{F}$ et $x_1, \ldots, x_n \in \mathbb{R}_+$ alors on obtient $\mathbb{E}(X) = \sum_{k=1}^n x_k \mathbb{P}(A_k)$ en vertu des deux premières propriétés. Si X est une variable aléatoire discrète positive, alors il existe une suite $(X_n)_{n\geq 1}$ croissante de v.a.r. du type précédent telles que $X = \lim_{n\to\infty} X_n$ d'où $\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$ par convergence monotone pour les séries. On admet le résultat au delà des v.a.r. discrètes. □

Exercice 3.19. *Montrer que si $X \geq 0$ et $\mathbb{E}(X) < \infty$ alors $\mathbb{P}(X < \infty) = 1$.*

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Si $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ est v.a.r. alors, en posant, $X_+ = \max(X, 0)$ et $X_- = \min(X, 0)$,

$$|X| = X_+ + X_- \quad \text{et} \quad X = X_+ - X_-.$$

Définition 3.20 (Espérance). *On dit qu'une v.a.r. $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ est* intégrable *lorsque $\mathbb{E}|X| = \mathbb{E}(X_+) + \mathbb{E}(X_-) < \infty$. On définit alors l'espérance de X en posant*

$$\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-).$$

On note $L^1(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des v.a.r. intégrables, et plus généralement $L^p(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des v.a.r. X telles que $|X|^p$ est intégrable.

L'espérance possède les propriétés fondamentales et immédiates suivantes :

1. **linéarité :** si $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $a, b \in \mathbb{R}$ alors $aX + bY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

2. **positivité :** si $X \in L_+(\Omega, \mathcal{F}, \mathbb{P})$ alors $\mathbb{E}(X) \geq 0$
3. **croissance :** si $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ avec $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$

L'inégalité triangulaire donne $|\mathbb{E}(X_+) - \mathbb{E}(X_-)| \leq \mathbb{E}(X_+) + \mathbb{E}(X_-)$ et donc

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$

Notons que $\mathbb{E}$ est une forme linéaire sur l'espace vectoriel $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Si X est une v.a.r. constante et égale à un réel c alors X est intégrable et $\mathbb{E}(X) = c$.

Définition 3.21 (Moments). *Si $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ avec $p \in \mathbb{N}^*$ alors on dit que $\mathbb{E}(X^p)$ est le* moment *d'ordre p de X. En particulier, $\mathbb{E}(X)$ est «le premier moment» de X.*

Remarque 3.22. *Si X est bornée, c'est-à-dire $\mathbb{P}(|X| \leq r) = 1$ pour un réel $r \in \mathbb{R}_+$, alors $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ pour tout $p \geq 1$ et X possède dans ce cas des moments de tout ordre.*

Remarque 3.23 (Caractérisation de la loi par une classe de fonctions tests). *La loi d'une v.a.r. X est caractérisée par la donnée de $\{\mathbb{E}(f(X)) : f \in \mathcal{F}\}$ pour une classe de fonctions $\mathcal{F}$ assez riche, comme par exemple les classes suivantes :*

- *Indicatrices d'intervalles (fonction de répartition !)*
- *Fonctions continues bornées $\mathbb{R} \to \mathbb{R}$ (approcher les indicatrices)*
- *Fonctions mesurables positives $\mathbb{R} \to \mathbb{R}$ (contient les indicatrices d'intervalles !)*
- *Fonctions de la forme $x \mapsto e^{itx}$ avec $t \in \mathbb{R}$ (transformée de Fourier ou fonc. carac.)*
- *Fonctions de la forme $x \mapsto s^x$ avec $s \in [0, 1]$ (fonction génératrice, si X discrète)*
- *Fonctions de la forme $x \mapsto e^{-tx}$ avec $t \geq 0$ (transformée de Laplace, si $X \geq 0$)*

Exercice 3.24 (Espérance des indicatrices). *Le principe d'inclusion-exclusion (théorème 2.17) découle via les propriétés de l'espérance de l'identité suivante :*

$$1 - \mathbf{1}_{\cup_{1 \leq i \leq r} A_i} = \prod_{1 \leq i \leq r} \mathbf{1}_{A_i^c} = \prod_{1 \leq i \leq r} (1 - \mathbf{1}_{A_i}) = \sum_{k=1}^{r} \sum_{1 \leq i_1 < \cdots < i_k \leq r} (-1)^k \mathbf{1}_{A_{i_1} \cap \cdots \cap A_{i_k}}.$$

Les inégalités de Boole-Bonferonni raffinent le principe d'inclusion-exclusion :

$$\mathbb{P}(\cup_{1 \leq i \leq r} A_i) - \sum_{k=1}^{m} (-1)^{k+1} S_k \quad \text{est} \quad \begin{cases} \geq 0 & \text{si } m \text{ impair} \\ \leq 0 & \text{si } m \text{ pair} \\ = 0 & \text{si } m = r \text{ (inclusion-exclusion !)} \end{cases}$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

En prenant $m = 1$ à gauche et $m = 2$ à droite on obtient le cas particulier (pour $r \geq 2$)

$$\underbrace{\mathbb{P}(A_1) + \cdots + \mathbb{P}(A_r)}_{S_1} \leq \mathbb{P}(\cup_{1 \leq i \leq r} A_i) \leq \underbrace{\mathbb{P}(A_1) + \cdots + \mathbb{P}(A_r)}_{S_1} - \underbrace{\sum_{i<j} \mathbb{P}(A_i \cap A_j)}_{S_2}.$$

Pour établir les inégalités de Boole-Boneferroni, on commence par observer que si

$$x_0 \leq \cdots \leq x_k \geq x_{k+1} \geq \cdots \geq x_r$$

sont des nombres réels vérifiant $\sum_{k=0}^{r}(-1)^k x_k = 0$ alors $\sum_{k=0}^{m}(-1)^k x_k \geq 0$ pour les m pairs et ≤ 0 pour les m impairs (ceci mérite une démonstration, qui est omise). Appliquée à la suite des coefficients binomiaux $x_0 = \binom{r}{0}, \ldots, x_r = \binom{r}{r}$, cette observation donne

$$\sum_{k=0}^{m}(-1)^k \binom{r}{k} \quad \text{est} \quad \begin{cases} \geq 0 & \text{si } m \text{ impair} \\ \leq 0 & \text{si } m \text{ pair} \\ = 0 & \text{si } m = r \text{ (formule du binôme !).} \end{cases}$$

À présent, si $r(\omega)$ désigne le nombre d'indices $j \in \{1, \ldots, r\}$ tels que $\omega \in A_j$ alors

$$\sum_{1 \leq i_1 < \cdots < i_k \leq r} \mathbf{1}_{A_{i_1}}(\omega) \cdots \mathbf{1}_{A_{i_k}}(\omega) = \binom{r(\omega)}{k}.$$

Le résultat désiré découle ensuite de la linéarité et de la positivité de l'espérance. On trouvera une application des inégalités de Boole-Boneferroni en fiabilité dans le livre de Delmas et Jourdain (pages 298–299).

3.3 Espérance des v.a.r. discrètes

Le théorème du transfert permet de calculer l'espérance de fonctions de v.a.r. discrètes. En particulier, il montre que $\mathbb{E}(\varphi(X))$ ne dépend que de la loi de X. Notons que si X est discrète alors $\varphi(X)$ aussi (pas forcément à valeurs dans $\mathbb{N}$ même si c'est le cas de X).

Théorème 3.25 (Espérance et formule du transfert pour les v.a.r. discrètes)**.** *Si X est une v.a.r. discrète à valeurs dans un ensemble au plus dénombrable E alors $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si la série $\sum_{x \in E} |x| \mathbb{P}(X = x)$ converge et on a alors*

$$\mathbb{E}(X) = \sum_{x \in E} x \mathbb{P}(X = x).$$

Plus généralement, pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$, on a $\varphi(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si la série $\sum_{x \in E} |\varphi(x)| \mathbb{P}(X = x)$ converge et on a alors la formule du transfert

$$\mathbb{E}(\varphi(X)) = \sum_{x \in E} \varphi(x) \mathbb{P}(X = x).$$

Démonstration. Quitte à numéroter les éléments de E, on peut supposer que $E = \mathbb{N}$. En écrivant $\varphi = \varphi_+ - \varphi_-$ on se ramène au cas où $\varphi \geq 0$. On a par convergence monotone

$$\mathbb{E}(\varphi(X)) = \lim_{n \to \infty} \mathbb{E}(\varphi(X) \mathbf{1}_{\{X \leq n\}})$$

et le résultat découle alors du fait que $\mathbb{E}(\varphi(X) \mathbf{1}_{\{X \leq n\}}) = \sum_{k=0}^{n} \varphi(k) \mathbb{P}(X = k)$. □

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Exercice 3.26. *Retrouver la formule du transfert pour les v.a.r. discrètes en partant de la formule* $\mathbb{E}(\varphi(X)) = \sum_z z\mathbb{P}(\varphi(X) = z)$. *De même, retrouver la linéarité de l'espérance pour les v.a.r. discrètes en partant de la formule* $\mathbb{E}(X + Y) = \sum_z z\mathbb{P}(X + Y = z)$.

Exercice 3.27. *Montrer que si* X *est une v.a.r. sur* $\mathbb{N}$ *intégrable alors*

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X \geq n).$$

Exercice 3.28 (Inégalité de Jensen). *Soit* $\varphi : \mathbb{R} \to \mathbb{R}$ *une fonction convexe et* X *une v.a.r. discrète telle que* X *et* $\varphi(X)$ *sont intégrables. On a alors*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

De plus, l'égalité est atteinte si $\mathbb{P}(X = c) = 1$ *pour une constante* c, *et réciproquement si* φ *est strictement convexe. Indication : se ramener au cas où* X *est discrète finie, utiliser le théorème du transfert et la définition de la convexité. Cas particuliers importants :* $\varphi(x) = |x|$, $\varphi(x) = x^2$, $\varphi(x) = |x|^p$ *avec* $p \geq 2$, $\varphi(x) = e^x$, $\varphi(x) = x\ln(x)$ *(si* $X \geq 0$*).*

Les fonctions génératrices sont utilisées pour les v.a.r. et lois discrètes sur $\mathbb{N}$. Pour les v.a.r. et lois continues, on utilise plutôt les transformées de Laplace ou les fonctions caractéristiques (hors programme). Le principe est le même dans tous les cas : on introduit une famille à un paramètre de fonctions mesurables $(f_s)_{s\in I}$, suffisamment riche pour que la connaissance de $(\mathbb{E}(f_s(X)))_{s\in I}$ caractérise complètement la loi de X.

Définition 3.29 (Fonction génératrice). *Si* P *est une loi de probabilité sur* $\mathbb{N}$ *alors sa fonction génératrice, notée* $g_P : [0, 1] \to \mathbb{R}$, *est définie pour tout* $s \in [0, 1]$ *par*

$$g_P(s) = \sum_{n=0}^{\infty} s^n P(\{n\}).$$

Si X *est une v.a.r. de loi* P *sur* $\mathbb{N}$ *alors on note* $g_X = g_P$ *de sorte que pour tout* $s \in [0, 1]$,

$$g_X(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n) = \mathbb{E}(s^X).$$

Remarque 3.30 (Transformée de Laplace). *On a* $t \in \mathbb{R}_+ \mapsto \mathbb{E}(e^{-tX}) = g_X(e^{-t})$.

Remarque 3.31 (Fonction caractéristique). *On peut définir* g_X *sur* $\{z \in \mathbb{C} : |z| \leq 1\}$. *La fonction caractéristique de* X *est* $\theta \in \mathbb{R}_+ \mapsto \mathbb{E}(e^{i\theta X}) = g_X(e^{i\theta})$, *c'est-à-dire* $g_X|_{\{z\in\mathbb{C}:|z|=1\}}$.

Théorème 3.32 (Fonctions génératrices). *Si* $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{N}$ *sont des v.a.r. alors :*

1. g_X *est croissante sur l'intervalle* $[0, 1]$, *avec* $g_X(0) = \mathbb{P}(X = 0)$ *et* $g_X(1) = 1$
2. g_X *est* $\mathcal{C}^\infty$ *sur* $]-1, 1[$ *et* $g_X^{(n)}(0) = \mathbb{P}(X = n)$ *pour tout* $n \in \mathbb{N}$
3. X *et* Y *ont même loi si et seulement si* $g_X = g_Y$
4. *si* $X(X - 1)\cdots(X - k + 1)$ *est intégrable* ($k \in \mathbb{N}^*$) *alors*

$$\mathbb{E}(X(X - 1)\cdots(X - k + 1)) = \lim_{s\nearrow 1} g_X^{(k)}(s).$$

L'expression $\mathbb{E}(X(X - 1)\cdots(X - k + 1))$ est appelée *moment factoriel* d'ordre k de X.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. Le 1. est immédiat. Le 2. découle du fait que le rayon de convergence de la série entière $\sum_{n=0}^{\infty} z^n \mathbb{P}(X = n)$ est supérieur ou égal à 1. Pour le 3., si $g_X = g_Y$ alors $\mathbb{P}(X = n) = g_X^{(n)}(0) = g_Y^{(n)}(0) = \mathbb{P}(Y = n)$ pour tout $n \in \mathbb{N}$ et la réciproque est évidente. Le 4. s'obtient en calculant $g_X^{(n)}(s)$ pour $s \in]0, 1[$ (dérivation sous le signe somme) puis en utilisant le théorème de convergence monotone (faire tendre s vers 1). □

Exemple 3.33 (Fonctions génératrices).

1. *Loi de Bernoulli :* $g(s) = 1 + (s-1)p$
2. *Loi binomiale :* $g(s) = (1 + (s-1)p)^n$
3. *Loi géométrique sur* $\mathbb{N}^*$ *:* $g(s) = sp/(1 - s(1-p))$
4. *Loi géométrique sur* $\mathbb{N}$ *:* $g(s) = p/(1 - s(1-p))$
5. *Loi de Poisson :* $g(s) = e^{\lambda(s-1)}$

Remarque 3.34 (Combinatoire). *Le n° nombre de Bell B_n compte le nombre de partitions d'un ensemble à n éléments. On a $B_0 = 1$, $B_1 = 1$, $B_2 = 2$, et $(B_n)_{n\geq 0}$ vérifie la récurrence*

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

qui se démontre de la manière suivante : pour choisir une partition de $\{1, \ldots, n+1\}$ on choisit le nombre k d'éléments qui n'appartiennent pas au même bloc que 1, puis ces k éléments parmi n, puis on partitionne ces k éléments avec les B_k possibilités. La formule de récurrence se réécrit de la manière suivante :

$$\frac{B_{n+1}}{n!} = \sum_{k_1+k_2=n} \frac{1}{k_1!} \frac{B_{k_2}}{k_2!}$$

ce qui donne l'identité des séries formelles en Z suivante :

$$\sum_{n=0}^{\infty} \frac{B_{n+1}}{n!} Z^n = \sum_{k_1=0}^{\infty} \frac{1}{k_1!} Z^{k_1} \sum_{k_2=0}^{\infty} \frac{B_{k_2}}{k_2!} Z^{k_2} = e^Z \sum_{n=0}^{\infty} \frac{B_n}{n!} Z^n,$$

qui s'écrit $G'(Z) = e^Z G(Z)$ où $G(Z) = \sum_{n=0}^{\infty} \frac{B_n}{n!} Z^n$, ce qui donne $G(Z) = e^{e^Z - 1}$. On reconnaît la fonction génératrice de la loi de Poisson de paramètre 1. On a donc

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

Notons par ailleurs que si $\left\{ {n \atop k} \right\}$ désigne le nombre de partitions à k blocs d'un ensemble à n éléments (nombre de Stirling de seconde espèce) alors

$$B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\}.$$

On dispose de la formule de récurrence

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\} \quad \text{avec conditions au bord} \quad \left\{ {n \atop 1} \right\} = 1 \quad \text{et} \quad \left\{ {n \atop n} \right\} = 1$$

car pour choisir une partition de $\{1, \ldots, n+1\}$ ayant k blocs il faut et il suffit soit de choisir une partition de $\{1, \ldots, n\}$ ayant $k-1$ blocs et de la compléter avec le bloc singleton $\{n+1\}$,

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

soit d'ajouter l'élément $n+1$ à l'un des k blocs d'une partition partition de $\{1,\dots,n\}$ ayant k blocs. Si X est une variable aléatoire de loi de Poisson de paramètre λ alors

$$\mathbb{E}(X^n) = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} \lambda^k \quad \text{en particulier} \quad \mathbb{E}(X^n) = B_n \quad \text{si } \lambda = 1.$$

On dispose également de la formule explicite suivante :

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^n$$

qui peut s'obtenir grâce au principe d'inclusion-exclusion en remarquant que le nombre de Stirling de seconde espèce est égal au nombre de surjections de $\{1,\dots,n\}$ dans $\{1,\dots,k\}$.

3.4 Espérance des v.a.r. à densité

Le théorème du transfert permet de calculer l'espérance de v.a.r. à densité et plus généralement l'espérance de fonctions de v.a.r. à densité. Le théorème du transfert montre en particulier que $\mathbb{E}(\varphi(X))$ ne dépend que de la loi de X via sa densité. Attention : si X est à densité, alors $\varphi(X)$ n'est pas forcément à densité.

Théorème 3.35 (Espérance et formule du transfert pour les v.a.r. à densité – Admis)**.** *Si X est une v.a.r. continue de densité f alors $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si la fonction $x \mapsto |x| f(x)$ est intégrable sur $\mathbb{R}$ et on a alors*

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\, dx.$$

Plus généralement, pour toute $\varphi : \mathbb{R} \to \mathbb{R}$ borélienne, on a $\varphi(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ ssi la fonction $x \mapsto |\varphi(x)| f(x)$ est intégrable sur $\mathbb{R}$ et on a alors la formule du transfert

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx.$$

Démonstration. Pour tout intervalle $I \subset \mathbb{R}$, on a , avec $\varphi = \mathbf{1}_I$, par définition de $\mathbb{E}$ et f,

$$\mathbb{E}(\varphi(X)) = \mathbb{E}(\mathbf{1}_I) = \mathbb{P}(X \in I) = \int_I f(x)\, dx = \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx.$$

On admet que cette formule reste valable lorsque I est un borélien de $\mathbb{R}$. La formule reste vraie par linéarité pour toute fonction φ étagée (i.e. constante sur un nombre fini de boréliens). Pour établir le résultat pour toute fonction φ borélienne, on se ramène tout d'abord au cas où $\varphi \geq 0$ en utilisant la décomposition $\varphi = \varphi_+ - \varphi_-$, puis on considère une suite croissante $(\varphi_n)_{n\geq 1}$ de fonctions positives étagées (constantes sur un nombre fini de boréliens) convergeant vers φ (existence admise) et on obtient par convergence monotone $\mathbb{E}(\varphi(X)) = \lim_{n\to\infty} \mathbb{E}(\varphi_n(X))$ et $\lim_{n\to\infty} \mathbb{E}(\varphi_n(X)) = \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx$. □

Exercice 3.36 (Moyenne)**.** *En utilisant le théorème du transfert, retrouver les formules pour la moyenne de la loi uniforme, exponentielle, et normale. La loi de Cauchy possède-t-elle une moyenne ? Une médiane ? Idem pour les lois de Pareto et de Student.*

Remarque 3.37 (Queues lourdes)**.** *Les lois sans espérance comme la loi de Cauchy ne sont pas des objets exotiques réservés aux contres exemples : de nombreux phénomènes naturels donnent des échantillons répartis en lois de puissance comme la loi de Cauchy, de Pareto, de Student. Notons que si X n'a pas de variance alors X^2 n'a pas d'espérance.*

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Exercice 3.38 (Calcul des moments). *En utilisant l'intégration par parties, montrer que les moments de la loi exponentielle de paramètre λ sont donnés pour tout $n \geq 1$ par*

$$\int_0^{\infty} x^n \lambda e^{-\lambda x}\, dx = \frac{n!}{\lambda^n}$$

(formule de récurrence $m_{n+1} = \lambda^{-1}(n+1)m_n$). Montrer que les moments d'ordre impairs de la loi normale standard $\mathcal{N}(0,1)$ sont nuls tandis que les moments pairs sont donnés par

$$\int_{-\infty}^{+\infty} x^{2n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\, dx = \prod_{k=1}^{n}(2k-1) = \frac{(2n)!}{2n(2n-2)\cdots 2} = \frac{(2n)!}{2^n n!}.$$

Montrer que les moments d'ordre impairs de la loi du demi cercle de densité

$$x \mapsto \frac{2}{\pi}\sqrt{1-x^2}\mathbf{1}_{[-1,1]}(x)$$

sont nuls tandis que les moments d'ordre pair sont les nombres de Catalan $\frac{1}{n+1}\binom{2n}{n}$.

Exercice 3.39. *Soit X une v.a.r. positive et intégrable, de densité f et de fonction de répartition F. Montrer en utilisant le théorème de convergence dominée que*

$$\lim_{r\to\infty} r\mathbb{P}(X > r) = \lim_{r\to\infty} \mathbb{E}(r\mathbf{1}_{]r,\infty[}(X)) = 0.$$

En déduire par intégration par parties basée sur $-(1-F)' = f$ sur un intervalle $[0, r]$ que

$$\mathbb{E}(X) = \lim_{r\to\infty} \int_0^r x f(x)\, dx = \int_0^{\infty} \mathbb{P}(X > x)\, dx.$$

Plus généralement, soit X une v.a.r. pas forcément positive, telle que $|X|^p$ est intégrable pour un réel $p \geq 1$. Montrer au moyen du théorème de Fubini-Tonelli que

$$\mathbb{E}(|X|^p) = p\int_0^{\infty} t^{p-1}\mathbb{P}(|X| > t)\, dt.$$

Cette identité est importante : elle relie moments et queue de distribution.

3.5 Variance

Rappelons que $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \{X : (\Omega, \mathcal{F}, \mathbb{R}) \to \mathbb{R}$ v.a.r. de carré intégrable$\}$.

Théorème 3.40 (Carré intégrable). *L'ensemble $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace vectoriel et si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ alors $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. En particulier, $L^2(\Omega, \mathcal{F}, \mathbb{P}) \subset L^1(\Omega, \mathcal{F}, \mathbb{P})$.*

Démonstration. On a $(X+Y)^2 \leq 2(X^2+Y^2)$ et donc $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace vectoriel. De plus, $XY = \frac{1}{2}((X+Y)^2 - X^2 - Y^2)$ et donc $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. □

Comme $\mathbb{E}(X^2) \geq 0$, l'application $(X, Y) \mapsto \mathbb{E}(XY)$ définit un produit scalaire sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$ et en particulier, on dispose de l'inégalité de Cauchy-Schwarz :

$$|\mathbb{E}(XY)| \leq \mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Définition 3.41 (Variance). *La variance de $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ est le nombre réel* ***positif***

$$\sigma^2(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

La variance de X représente la moyenne du carré des écarts à la moyenne. Pour obtenir un nombre de même dimension que X, on définit l'*écart-type* de X par

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

Comme $\sigma^2(X) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2)$, on obtient la formule de König

$$\sigma^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

On a $\sigma^2(X) = 0$ si et seulement si $\mathbb{P}(X = \mathbb{E}(X)) = 1$ (X est presque sûrement constante).

Remarque 3.42 (Moindres carrés). *Si $X \in \mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ alors*

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \inf_{m \in \mathbb{R}} \mathbb{E}((X - m)^2).$$

L'inf est atteint en $m = \mathbb{E}(X)$. Ainsi $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ est la distance des moindres carrés (i.e. L^2) de X au sous-espace vectoriel de $\mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ des v.a.r. constantes. En d'autres termes, $\mathbb{E}(X)$ est la constante la plus proche de X au sens L^2 et cette distance vaut σ_X.

Exercice 3.43. *Recalculer la moyenne et la variance pour les lois classiques (table 3.1).*

Remarque 3.44. *Si X est une v.a.r. positive de carré intégrable alors par l'inégalité de Cauchy-Schwarz, $\mathbb{E}(X) = \mathbb{E}(X\mathbf{1}_{\{X>0\}}) \leq \mathbb{E}(X^2)^{1/2}\mathbb{P}(X > 0)^{1/2}$ d'où,*

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)}.$$

3.6 Inégalités de Markov et de Bienaymé-Tchebychev

Comme $\mathbb{R}$ est réunion dénombrable de compacts, si X est une v.a.r. alors pour tout $\varepsilon > 0$ il existe un compact $K \subset \mathbb{R}$ tel que $\mathbb{P}(X \notin K) \leq \varepsilon$, et on dit que la loi de X est *tendue*. Les inégalités de Markov et de Bienaymé-Tchebychev ci-dessous permettent de mieux quantifier la propriété de tension lorsque X possède des moments finis.

Théorème 3.45 (Inégalité de Markov). *Si $0 \leq X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ alors pour tout $r > 0$*

$$\mathbb{P}(X \geq r) \leq \frac{\mathbb{E}(X)}{r}.$$

L'inégalité de Markov n'a pas d'intérêt quand $r \leq \mathbb{E}(X)$ (la borne est alors ≥ 1). L'inégalité de Markov affirme qu'une v.a.r. $X \geq 0$ intégrable est toujours concentrée autour de 0. Le cas extrême $\mathbb{E}(X) = 0$ entraîne $\mathbb{P}(X = 0) = 1$ (masse de Dirac en 0). Si par exemple $\mathbb{E}(X) = 1$ alors X dépasse 100 avec une probabilité inférieure ou égale à 1%.

Démonstration. Découle de la croissance de l'espérance utilisée avec $r\mathbf{1}_{\{X \geq r\}} \leq X$. □

Théorème 3.46 (Bienaymé-Tchebychev). *Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ alors pour tout $r > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq r) \leq \frac{\sigma^2(X)}{r^2}.$$

Démonstration. L'inégalité de Markov (théorème 3.45) pour la v.a.r. $(X - \mathbb{E}(X))^2$ donne

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq r) = \mathbb{P}((X - \mathbb{E}(X))^2 \geq r^2) \leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{r^2} = \frac{\sigma^2(X)}{r^2}.$$

□

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Loi	Support	Poids ou densité	Moyenne	Variance
Bernoulli	$\{0,1\}$	$\mathbb{P}(X=1)=p$	p	$p(1-p)$
Rademacher	$\{-1,1\}$	$\mathbb{P}(X=1)=p$	$2p-1$	$p(1-p)$
Binomiale	$\{0,1,\ldots,n\}$	$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Poisson	$\mathbb{N}$	$\mathbb{P}(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ
Hypergéom.	$\{0,1,\ldots,n\}$	$\mathbb{P}(X=k)=\frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}$	$n\frac{N_1}{N}$	$\frac{nN_1N_2(N-n)}{N^2(N-1)}$
Géom.	$\mathbb{N}^*$	$\mathbb{P}(X=k)=(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{p}{(1-p)^2}$
Géom.	$\mathbb{N}$	$\mathbb{P}(X=k)=(1-p)^kp$	$\frac{1-p}{p}$	$\frac{p}{(1-p)^2}$
Zipf	$\mathbb{N}^*$	$\mathbb{P}(X=k)=\frac{1}{k^s}$	$\zeta(s)^{(s>2)}$	$\zeta(s-2)-\zeta(s)^{2\,(s>3)}$
Uniforme	$\{1,\ldots,n\}$	$\mathbb{P}(X=k)=\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Uniforme	$[a,b]\subset\mathbb{R}$	$x\mapsto\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$
Cauchy	$\mathbb{R}$	$x\mapsto\frac{1}{\pi(1+x^2)}$	non	non
Student	$\mathbb{R}$	$x\mapsto\frac{\Gamma(\frac{1}{2}(a+1))}{\sqrt{a\pi}\Gamma(\frac{1}{2}a)}\left(1+\frac{x^2}{a}\right)^{-\frac{1}{2}(a+1)}$	$0^{\,(a>1)}$	$\frac{a}{a-2}^{(a>2)}$
Pareto	$[1,\infty[$	$x\mapsto\frac{n}{x^{n+1}}\mathbf{1}_{[1,\infty[}(x)$	$\frac{n}{n-1}^{(n>1)}$	$\frac{n}{(n-1)^2(n-2)}^{(n>2)}$
Expo.	$\mathbb{R}_+$	$x\mapsto\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Laplace	$\mathbb{R}$	$x\mapsto\frac{\lambda}{2}e^{-\lambda\lvert x\rvert}$	0	$\frac{1}{\lambda^2}$
Beta	$[0,1]\subset\mathbb{R}$	$x\mapsto\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\mathbf{1}_{[0,1]}(x)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
Gamma	$\mathbb{R}_+$	$x\mapsto\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$
Normale	$\mathbb{R}$	$x\mapsto\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-m)^2/(2\sigma^2)}$	m	σ^2
χ^2	$\mathbb{R}_+$	$x\mapsto\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$	n	$2n$

Table 3.1 – Quelques lois univariées sur $\mathbb{N}$ ou $\mathbb{R}$. La loi de Cauchy est une loi de Student avec $a=1$. La loi exponentielle est une loi Gamma avec $a=1$. La loi du $\chi^2(n)$ est une loi Gamma avec $a=n/2$. La loi uniforme continue est une loi Beta avec $a=b=1$. La loi Beta avec $a=b=1/2$ est la loi de l'arc-sinus et et du demi-cercle si $a=b=3/2$.

L'inégalité n'a pas d'intérêt lorsque $r\leq\sigma(X)$ car la borne est alors plus grande que 1.

Plus généralement, soit $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ une fonction mesurable croissante telle que $\varphi(r)>0$ pour tout $r>0$. Si $\varphi(|X-\mathbb{E}(X)|)$ est intégrable alors pour tout $r>0$

$$\mathbb{P}(|X-\mathbb{E}(X)|\geq r)\leq\frac{\mathbb{E}(\varphi(|X-\mathbb{E}(X)|))}{\varphi(r)}.$$

Pour $\varphi(r)=r^2$ on retrouve l'inégalité de Bienaymé-Tchebychev. Autres exemples classiques : $\varphi(r)=r^p$ avec $p\geq 1$, et $\varphi(r)=\exp(r)$. La morale est la suivante : plus X est intégrable plus on peut affirmer qu'elle est concentrée autour de sa moyenne.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Chapitre 4

Vecteurs aléatoires

On rappelle qu'un pavé de $\mathbb{R}^d$ est un produit de d intervalles de $\mathbb{R}$.

Définition 4.1 (Vecteur aléatoire). *Un vecteur aléatoire $X = (X_1, \ldots, X_d)$ de $\mathbb{R}^d$ est une suite $X_1, \ldots, X_d$ de variables aléatoires réelles définies sur un même espace $(\Omega, \mathcal{F}, \mathbb{P})$. La loi de X est la mesure de probabilité $\mathbb{P}_X$ sur $\mathbb{R}^d$ définie pour tout pavé $I_1 \times \cdots \times I_d$ par*

$$\mathbb{P}_X(I_1 \times \cdots \times I_d) = \mathbb{P}(X_1 \in I_1, \ldots, X_d \in I_d) = \mathbb{P}(X \in I_1 \times \cdots \times I_d).$$

Les lois des v.a.r. $X_1, \ldots, X_d$ sont les lois marginales *du vecteur aléatoire X.*

Le vecteur aléatoire X de $\mathbb{R}^d$ est discret lorsque $X(\Omega)$ est au plus dénombrable, et sa loi est alors entièrement déterminée par la donnée pour tout $(x_1, \ldots, x_d) \in X(\Omega)$ de

$$\mathbb{P}(X_1 = x_1, \ldots, X_d = x_d).$$

Les lois marginales sont des v.a.r. discrètes et leur loi s'obtient en sommant par rapport à toutes les autres variables. Par exemple, la loi de la v.a.r. X_1 est donnée par

$$\mathbb{P}(X_1 = x_1) = \sum_{x_2 \in X_2(\Omega), \ldots, x_d \in X_d(\Omega)} \mathbb{P}(X_1 = x_1, \ldots, X_d = x_d).$$

On dit qu'une fonction $f : \mathbb{R}^d \to \mathbb{R}$ est une *densité* de probabilité lorsque

$$f \geq 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_d)\, dx_1 \cdots dx_d = 1.$$

On dit que le vecteur aléatoire X admet pour densité f lorsque pour tout pavé $I_1 \times \cdots \times I_d$,

$$\mathbb{P}(X \in I_1 \times \cdots \times I_d) = \int_{I_1} \cdots \int_{I_d} f(x_1, \ldots, x_d)\, dx_1 \cdots dx_d.$$

La densité d'un vecteur n'est pas unique car on peut légèrement la modifier. Les lois marginales sont également à densité et leur densité s'obtient en intégrant f par rapport à toutes les autres variables (prendre $I_j = \mathbb{R}$ pour $j \neq i$). Par exemple, la densité de X_1 est

$$x_1 \in \mathbb{R} \mapsto \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_d)\, dx_2 \ldots dx_d.$$

Le calcul peut être mené par intégrations successives grâce au théorème de Fubini-Tonelli.

$\boxed{\text{Le théorème de Fubini-Tonelli est fondamental et simple : ne pas en avoir peur !}}$

Copyrighted material

Exemple 4.2 (Loi multinomiale). *Soit $p_1, \dots, p_d \in [0,1]$ tels que $p_1 + \cdots + p_d = n$. On dit que le vecteur aléatoire $X = (X_1, \dots, X_d)$ suit la loi multinomiale de taille $n \in \mathbb{N}^*$ et de paramètres $p_1, \dots, p_d$ lorsque pour tout $(n_1, \dots, n_d) \in \mathbb{N}^d$ tel que $n_1 + \cdots + n_d = n$,*

$$\mathbb{P}(X_1 = n_1, \dots, X_d = n_d) = \frac{n!}{n_1! \cdots n_d!} p_1^{n_1} \cdots p_d^{n_d}.$$

Cette loi est associée à la formule du multinome

$$\sum_{\substack{(n_1,\dots,n_d)\in\mathbb{N}^d \\ n_1+\cdots+n_d=n}} \frac{n!}{n_1! \cdots n_d!} p_1^{n_1} \cdots p_d^{n_d} = (p_1 + \cdots + p_d)^n = 1.$$

Elle modélise le jet de n dés discernables à d faces, ou encore n tirages avec remise dans une urne contenant d boules numérotées (donc discernables). On peut établir que si $\varnothing \neq I \subset \{1, \dots, d\}$ alors la v.a.r. $\sum_{i \in I} X_i$ suit la binomiale $\mathrm{Binom}(n, \sum_{i \in I} p_i)$. En particulier, pour tout $1 \leq i \leq d$ la v.a.r. X_i suit la loi binomiale $\mathrm{Binom}(n, p_i)$. Plus généralement, si $I_1, \dots, I_r$ est une partition de $\{1, \dots, d\}$ alors $(\sum_{i \in I_1} X_i, \dots, \sum_{i \in I_r} X_i)$ suit la loi multinomiale de taille r et de paramètre $(\sum_{i \in I_1} p_i, \dots, \sum_{i \in I_r} p_i)$.

Exemple 4.3 (Loi hypergéométrique multitypes). *Considérons des entiers $d \geq 2$, $1 \leq n \leq N$, et $N_1, \dots, N_d \geq 1$ avec $N_1 + \cdots + N_d = N$. La loi hypergéométrique multitypes $\mathrm{HyperGeom}(N_1, \dots, N_d, n)$ permet de modéliser le résultat $(X_1, \dots, X_d)$ du tirage sans remise dans une population finie composée de d types d'individus. Elle est donnée pour tout $n = (n_1, \dots, n_d) \in \mathbb{N}^d$ tel que $n_1 + \cdots + n_d = n$ et $n_1 \leq N_1, \dots, n_d \leq N_d$ par*

$$\mathbb{P}(X_1 = n_1, \dots, X_d = n_d) = \frac{\binom{N_1}{n_1} \cdots \binom{N_d}{n_d}}{\binom{N}{n}}$$

(voir l'exemple 2.8). Pour $1 \leq i \leq d$, la composante X_i suit la loi hypergéométrique à deux types $\mathrm{HyperGeom}(N_i, N - N_i, n)$ de l'exemple 3.8. Lorsque $N_1, \dots, N_d \to \infty$ avec

$$(N_i/N, \dots, N_d/N) \to (p_1, \dots, p_d)$$

alors on retrouve la loi multinomiale de taille n et de paramètre $(p_1, \dots, p_d)$. Notons que si $I_1, \dots, I_r$ est une partition de $\{1, \dots, d\}$ alors $(\sum_{i \in I_1} X_i, \dots, \sum_{i \in I_r} X_i)$ suit la loi

$$\mathrm{HyperGeom}(\sum_{i \in I_1} N_i, \dots, \sum_{i \in I_r} N_i, n).$$

Cette propriété d'autosimilarité par contraction est l'analogue de celle de la loi multinomiale.

Exemple 4.4 (Loi normale ou loi de Gauss ou loi gaussienne). *Soit $m \in \mathbb{R}^d$ et Σ une matrices symétrique $d \times d$ dont les valeurs propres sont strictement positives. On dit que le vecteur aléatoire $X = (X_1, \dots, X_d)$ suit la normale de moyenne m et de matrice de covariance Σ, et on note $X \sim \mathcal{N}(m, \Sigma)$ lorsque X a pour densité*

$$x \in \mathbb{R}^d \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}\langle \Sigma^{-1} x, x \rangle\right).$$

La loi est qualifiée de standard *lorsque $m = 0$ et $\Sigma = I_d$.*

Théorème 4.5 (Théorème du transfert pour les vecteurs aléatoires – Admis). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ et $\varphi : \mathbb{R}^d \to \mathbb{R}$ une fonction borélienne.*

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

1. *Si X est discret et si la série $\sum_{x\in X(\Omega)} |\varphi(x)|\mathbb{P}(X=x)$ converge alors*

$$\mathbb{E}(\varphi(X_1,\dots,X_d)) = \sum_{(x_1,\dots,x_d)\in\Omega(X)} \varphi(x_1,\dots,x_d)\mathbb{P}(X_1=x_1,\dots,X_d=x_d).$$

2. *Si X a pour densité f et si $x\in\mathbb{R}^d \mapsto |\varphi(x)|f(x)$ est intégrable alors*

$$\mathbb{E}(\varphi(X_1,\dots,X_d)) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \varphi(x_1,\dots,x_d)\, f(x_1,\dots,x_d)\, dx_1\cdots dx_d.$$

Démonstration. Similaire à celle pour les v.a.r. □

Exercice 4.6 (Linéarité de l'espérance). *Soit (X,Y) un vecteur aléatoire de $\mathbb{R}^2$ de densité f. En utilisant le théorème du transfert pour les vecteurs aléatoires, retrouver la propriété de linéarité de l'espérance : $\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y)$.*

4.1 Indépendance et covariance

Définition 4.7 (Indépendance). *Si $X=(X_1,\dots,X_d)$ est un vecteur aléatoire de $\mathbb{R}^d$ alors on dit que les v.a.r. $X_1,\dots,X_d$ sont* indépendantes *lorsque pour tous intervalles $I_1,\dots,I_d$ de $\mathbb{R}$ les évènements $\{X_1\in I_d\},\dots,\{X_d\in I_d\}$ sont indépendants.*

Si par exemple A et B sont deux évènements alors les v.a.r. booléennes $\mathbf{1}_A$ et $\mathbf{1}_B$ sont indépendantes si et seulement si A et B sont indépendants. Une variable aléatoire constante est proportionnelle à $\mathbf{1}_\Omega$ et est donc indépendante de toutes les autres v.a.r.

Exercice 4.8 (De la loi de Poisson à la loi multinomiale). *Montrer que si $X_1,\dots,X_d$ sont des v.a.r. indépendantes de loi de Poisson $\mathrm{Poi}(\lambda_1),\dots,\mathrm{Poi}(\lambda_d)$, alors le vecteur aléatoire*

$$\left(\frac{X_1}{X_1+\cdots+X_d},\dots,\frac{X_d}{X_1+\cdots+X_d}\right)$$

suit la loi multinomiale de taille n et de paramètre

$$(p_1,\dots,p_d)=\left(\frac{\lambda_1}{\lambda_1+\cdots+\lambda_d},\dots,\frac{\lambda_d}{\lambda_1+\cdots+\lambda_d}\right).$$

Montrer qu'il est indépendant de la v.a.r. de Poisson $X_1+\cdots+X_d$.

Théorème 4.9 (Espérance et indépendance – Admis). *Si $X=(X_1,\dots,X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ alors $X_1,\dots,X_d$ sont indépendantes si et seulement si pour toutes fonctions boréliennes positives $\varphi_1,\dots,\varphi_d:\mathbb{R}\to\mathbb{R}_+$, on a, dans $\mathbb{R}_+\cup\{+\infty\}$,*

$$\mathbb{E}\left(\prod_{i=1}^d \varphi_i(X_i)\right) = \prod_{i=1}^d \mathbb{E}(\varphi_i(X_i)).$$

De plus, si $X_1,\dots,X_d$ sont indépendantes alors pour toutes fonctions $\varphi_1,\dots,\varphi_d:\mathbb{R}\to\mathbb{R}$ boréliennes vérifiant $\varphi_1(X_1),\dots,\varphi_d(X_d)\in L^1(\Omega,\mathcal{F},\mathbb{P})$, on a

$$\prod_{i=1}^d \varphi_i(X_i)\in L^1(\Omega,\mathcal{F},\mathbb{P}) \quad \textit{et} \quad \mathbb{E}\left(\prod_{i=1}^d \varphi_i(X_i)\right) = \prod_{i=1}^d \mathbb{E}(\varphi_i(X_i)).$$

En particulier $\mathbb{E}(X_1\cdots X_d)=\mathbb{E}(X_1)\cdots\mathbb{E}(X_d)$ si $X_1,\dots,X_d\in L^1(\Omega,\mathcal{F},\mathbb{P})$.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. Si la première propriété est vraie, alors son application à des fonctions de la forme $\varphi_i = \mathbf{1}_{I_i}$ où I_i est un intervalle de $\mathbb{R}$ fournit l'indépendance de $X_1, \ldots, X_d$. Réciproquement, on procède par approximation, linéarité de l'espérance, et convergence monotone à partir de ces fonctions élémentaires. Pour établir la seconde propriété, on utilise la $\varphi_i = (\varphi_i)_+ - (\varphi_i)_-$, la linéarité de l'espérance, et la première propriété. □

Théorème 4.10 (Indépendance et structure produit des densités – Admis)**.**

1. *Si les v.a.r. $X_1, \ldots, X_d$ sont indépendantes de densités $f_1, \ldots, f_d$ alors le vecteur $X = (X_1, \ldots, X_d)$ admet la densité $x \mapsto (f_1 \otimes \cdots \otimes f_d)(x) = f_1(x_1) \cdots f_d(x_d)$*
2. *Les composantes $X_1, \ldots, X_d$ d'un vecteur aléatoire X de $\mathbb{R}^d$ de densité f de marginales $f_1, \ldots, f_d$ sont indépendantes ssi X admet aussi $f_1 \otimes \cdots \otimes f_d$ comme densité.*

Démonstration. Découle de la définition de l'indépendance et du théorème de Fubini-Tonelli. □

Exercice 4.11 (Loi de Cauchy)**.** *Soient X et Y des v.a.r. indépendantes de loi normale $\mathcal{N}(0,1)$. En utilisant le théorème du transfert et un changement de variable, montrer que $C = X/Y$ suit la loi de Cauchy. En déduire que si C est de Cauchy alors $1/C$ l'est aussi.*

Remarque 4.12 (Lois normales)**.** *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ de loi normale $\mathcal{N}(m, \Sigma)$ où Σ est inversible. Si Σ est diagonale alors la densité de X est produit et donc $X_1, \ldots, X_d$ sont indépendantes. Réciproquement, si les $X_1, \ldots, X_n$ sont indépendantes alors, comme nous le verrons plus loin, la matrice de covariance Σ est diagonale car $\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = 0$ si $i \neq j$.*

Remarque 4.13 (Simulation de la loi normale)**.** *La fonction de répartition de la loi normale ne possède pas d'expression explicite ce qui ne rend pas commode l'usage de la méthode d'inversion pour sa simulation. On peut cependant utiliser l'agorithme polaire de Box-Muller : soit (X, Y) un vecteur aléatoire de $\mathbb{R}^2$ de coordonnées polaires (r, θ). Alors X et Y sont indépendantes de loi $\mathcal{N}(0,1)$ si et seulement si r et θ sont indépendantes avec $r^2 \sim \mathrm{Exp}(1/2) = \mathrm{Gamma}(1, 1/2) = \chi^2(2)$ et $\theta \sim \mathrm{Unif}([0, 2\pi])$. En effet :*

$$\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\, dxdy = r e^{-\frac{r^2}{2}} \mathbf{1}_{\mathbb{R}_+}(r) \frac{1}{2\pi} \mathbf{1}_{[0,2\pi]}(\theta) dr d\theta.$$

Cette méthode fournit d'emblée deux réalisations indépendantes de $\mathcal{N}(0,1)$ c'est-à-dire une réalisation de $\mathcal{N}(0, I_2)$. Pour des raisons de performance et de précision, certains logiciels utilisent plutôt une méthode de discrétisation-rejet (algorithme du Ziggurat de Marsaglia).

Théorème 4.14 (Indépendance et convolution pour le cas à densité)**.** *Si X, Y sont deux variables aléatoires indépendantes de densités f et g alors $X + Y$ admet la densité*

$$z \in \mathbb{R} \mapsto (f * g)(z) = \int_{-\infty}^{+\infty} f(z-y) g(y)\, dy.$$

Démonstration. Comme X et Y sont indépendantes, le couple (X, Y) admet pour densité la fonction produit $(x, y) \mapsto f(x)g(y)$. Le théorème du transfert pour le couple (X, Y) donne alors pour tout $t \in \mathbb{R}$,

$$\mathbb{P}(X + Y \leq t) = \iint \mathbf{1}_{\{(x,y): x+y \leq t\}} f(x) g(y)\, dxdy.$$

En effectuant le changement de variable $(x, y) \mapsto (z, y)$ avec $z = x + y$ il vient

$$\mathbb{P}(X + Y \leq t) = \iint \mathbf{1}_{\{z \leq t\}} f(z-y) g(y)\, dydz.$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Le théorème de Fubini-Tonelli donne à présent

$$\mathbb{P}(X+Y \leq t) = \int_{-\infty}^{t} \left(\int_{-\infty}^{+\infty} f(z-y)g(y)\,dy \right) dz.$$

□

Exercice 4.15 (Propriétés des lois normales). *En utilisant le produit de convolution, montrer que si X_1 et X_2 sont deux v.a.r. indépendantes de loi normales $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ alors $X+Y$ suit la loi normale $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$.*

Exercice 4.16 (Propriétés des lois normales). *Soit Σ une matrice $d \times d$ symétrique dont toutes les valeurs propres sont strictement positive. Soit A une matrice telle que $\Sigma = AA^\top$ (par exemple via le théorème spectral ou via la décomposition de Cholesky). Établir au moyen du théorème du transfert et d'un changement de variable que si $Z \sim \mathcal{N}(0, I_d)$ alors $AZ+m \sim \mathcal{N}(m, \Sigma)$. En déduire que si $X \sim \mathcal{N}(m, \Sigma)$ et $v \in \mathbb{R}^d$ alors $\langle X, v\rangle \sim \mathcal{N}(\langle m, v\rangle, \langle v, \Sigma v\rangle)$ et en particulier $X_i \sim \mathcal{N}(m_i, \Sigma_{i,i})$ pour tout $1 \leq i \leq d$. Ainsi, les lois marginales d'un vecteur aléatoire de loi normale sont toutes de loi normale.*

Exercice 4.17 (Propriétés des lois normales). *Montrer que si X et Y sont deux v.a.r. indépendantes avec X de loi de Rademacher de paramètre $1/2$ et Y de loi normale $\mathcal{N}(0,1)$ alors la v.a.r. XY suit la loi normale $\mathcal{N}(0,1)$ tandis que la v.a.r. $XY+Y$ ne suit pas la loi normale. En déduire que le vecteur aléatoire (XY, Y) de $\mathbb{R}^2$ a des lois marginales normales mais ne suit pas une loi normale sur $\mathbb{R}^2$. Les fonctions caractéristiques (hors programme) permettent d'établir un résultat positif dans cet esprit en allant au delà des simples lois marginales : si Z est un vecteur aléatoire de $\mathbb{R}^d$ tel que $v \cdot Z = v_1 Z_1 + \cdots + v_d Z_d$ suit une loi normale sur $\mathbb{R}$ pour tout $v \in \mathbb{R}^d$ alors Z est suit une loi normale sur $\mathbb{R}^d$.*

Définition 4.18 (Covariance). *La covariance de $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ est définie par*

$$\mathrm{Cov}(X,Y) = \mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y)) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

La v.a.r. $(X-\mathbb{E}X)(Y-\mathbb{E}Y)$ est intégrable car grâce à l'inégalité de Cauchy-Schwarz,

$$\mathbb{E}(|(X-\mathbb{E}X)(Y-\mathbb{E}Y)|) \leq \sigma(X)\sigma(Y).$$

On a toujours $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ et $\mathrm{Cov}(X,X) = \sigma^2(X)$ et la formule bilinéaire

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y) + 2\mathrm{Cov}(X,Y).$$

Si X, Y sont indépendantes alors $\mathrm{Cov}(X,Y) = 0$ (on dit que X, Y sont non corrélées) et

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y).$$

La réciproque est fausse : $\mathrm{Cov}(X,Y) = 0$ n'implique pas que X et Y sont indépendantes (on admet que la réciproque est cependant vraie si (X,Y) suit une loi gaussienne de $\mathbb{R}^2$). Contre exemple : $\mathrm{Cov}(U, U^2) = \mathbb{E}(U^3) - \mathbb{E}(U)\mathbb{E}(U^2) = 0$ si U est uniforme sur $[-1,1]$.

Exercice 4.19. *Soit X un vecteur aléatoire de $\mathbb{R}^d$ de loi normale $X \sim \mathcal{N}(m, \Sigma^2)$. Montrer que $\mathbb{E}(X_i) = m_i$ pour tout $1 \leq i \leq d$ et $\mathrm{Cov}(X_i, X_j) = \Sigma_{i,j}$ pour tout $1 \leq i, j \leq d$.*

Remarque 4.20 (Erreur fréquente). *Si $X_1, \ldots, X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ alors*

$$\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

par linéarité de l'espérance (nul besoin d'indépendance). Si $X_1, \ldots, X_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ alors

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

*Si maintenant $X_1, \ldots, X_n$ sont non corrélées (par exemple **indépendantes**) alors*

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n).$$

Mais la variance est quadratique, et non pas linéaire : $\sigma^2(aX + b) = a^2\sigma^2(X)$.

Définition 4.21 (Coefficient de corrélation de Pearson). *Si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ avec $\sigma^2(X) > 0$ et $\sigma^2(Y) > 0$ alors on appelle* corrélation *de X et Y la quantité*

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\sigma^2(X)}\sqrt{\sigma^2(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Le coefficient de corrélation mesure en quelque sorte la dépendance linéaire. En effet, l'inégalité de Cauchy-Schwarz et ses cas d'égalité permet d'établir les propriétés suivantes :

1. $-1 \leq \rho(X, Y) \leq 1$
2. $\rho(X, Y) = 1$ si et seulement si $\mathbb{P}(X = aY + b) = 1$ pour des réels $a > 0$ et b
3. $\rho(X, Y) = -1$ si et seulement si $\mathbb{P}(X = aY + b) = 1$ pour des réels $a < 0$ et b

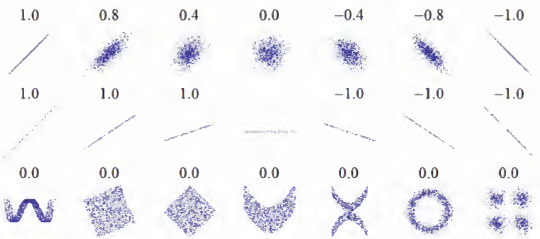


Figure 4.1 – Relation entre corrélation et dépendance linéaire (source : Wikipédia).

4.2 Matrice de covariance

Définition 4.22 (Vecteur moyenne et matrice de covariance). *Soit X un vecteur colonne aléatoire de $\mathbb{R}^d$. Si ses composantes sont intégrables alors on définit son vecteur moyenne*

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d))^\top.$$

Si ses composantes sont de carré intégrable alors on définit sa matrice de covariance $\Sigma(X)$ comme étant la matrice symétrique $d \times d$ donnée pour tous $1 \leq i, j \leq d$ par

$$\Sigma(X)_{i,j} = \mathrm{Cov}(X_i, X_j).$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

On dit que X est centré si $\mathbb{E}[X] = 0$. La diagonale de $\Sigma(X)$ est constituée des variances des composantes de X. Si les composantes de X sont indépendantes alors $\Sigma(X)$ est diagonale. La réciproque est fausse en général, mais vraie pour les vecteurs gaussiens.

Théorème 4.23 (Structure des matrices de covariance). *La matrice de covariance d'un vecteur aléatoire est toujours symétrique et semi-définie positive.*

Démonstration. Si v est un vecteur colonne de $\mathbb{R}^d$, alors la matrice $vv^\top$ est symétrique, semi-définie positive, de rang 1. Ses valeurs propres sont 0 et $v^\top v = \|v\|_2^2$. La matrice aléatoire $(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top$ est symétrique, semi-définie positive, de rang 1. Par conséquent, son espérance Σ est également symétrique. Le fait que Σ soit semi-définie positive découle de la linéarité de l'espérance, car si u est un vecteur colonne de $\mathbb{R}^d$,

$$u^\top \Sigma u = u^\top \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top]u = \mathbb{E}[u^\top (X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top u] \geq 0.$$

En revanche, Σ peut être de rang quelconque entre 1 et d, bien que $(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top$ soit de rang 1 (il faut concevoir l'espérance comme une combinaison convexe infinie). □

L'ensemble des matrices $d \times d$ symétriques semi-définies positives est un cône convexe fermé : si A et B en sont deux éléments, alors pour tous réels $\lambda, \mu \geq 0$, la matrice $\lambda A + \mu B$ l'est également. La frontière de ce cône est constituée par les éléments du cône qui ne sont pas de plein rang. L'intérieur du cône est le cône convexe ouvert des matrices symétriques définies positives de dimension $d \times d$.

Théorème 4.24 (Racines carrées matricielles). *Toute matrice symétrique semi-définie positive Σ de dimension $d \times d$ s'écrit $\Sigma = AA^\top$ où A est une matrice de dimension $d \times d$. Une telle matrice A, appelée* racine carrée *de Σ, n'est pas unique en général.*

Démonstration. Le théorème spectral fournit une matrice diagonale D et une matrice orthogonale P telles que $\Sigma = PDP^\top$. De plus, $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_d)$ où $\{\lambda_1, \ldots, \lambda_d\} \subset \mathbb{R}_+^d$ est le spectre de Σ. Ainsi, $\Sigma = AA^\top$ où $A = P\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$. Un autre choix possible est $A = P\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})P^\top$, qui fournit une racine carrée symétrique et semi-définie positive. Dans les deux cas, les matrices Σ et A ont le même rang. Alternativement, la décomposition de Cholesky fournit une matrice triangulaire inférieure A à diagonale positive ou nulle qui vérifie $AA^\top = \Sigma$. Une telle matrice se calcule par un algorithme récursif simple et explicite. En effet, l'équation $AA^\top = \Sigma$ est équivalente au système d'équations suivant : pour tous $1 \leq i \leq j \leq d$

$$A_{i,i}A_{j,i} = \Sigma_{i,j} - \sum_{k=1}^{i-1} A_{i,k}A_{j,k}.$$

La diagonale de Σ est positive ou nulle, strictement positive lorsque Σ est inversible. Dans ce dernier cas, il existe une unique matrice A triangulaire inférieure à diagonale strictement positive telle que $AA^\top = \Sigma$, et la décomposition de Cholesky constitue alors un cas particulier de la décomposition LU des matrices inversibles. □

Théorème 4.25 (Transformations linéaires). *Si X est un vecteur colonne aléatoire de $\mathbb{R}^d$ à composantes de carré intégrable et si A est une matrice $n \times d$, alors le vecteur aléatoire AX de $\mathbb{R}^n$ a pour vecteur moyenne $A\mathbb{E}(X)$ et pour matrice de covariance $A\Sigma(X)A^\top$.*

Démonstration. Par linéarité de l'espérance, on a $\mathbb{E}(AX) = A\mathbb{E}(X)$ et, lorsque $\mathbb{E}(X) = 0$,

$$\mathbb{E}((AX)(AX)^\top) = \mathbb{E}(AXX^\top A^\top) = A\mathbb{E}(XX^\top)A^\top = A\Sigma(X)A^\top.$$

□

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Théorème 4.26 (Généricité des matrices de covariance). *Toute matrice $d \times d$ Σ symétrique semi-définie positive est la matrice de covariance d'un vecteur aléatoire de $\mathbb{R}^d$.*

Démonstration. Soit A une racine carrée matricielle de Σ et X un vecteur aléatoire de $\mathbb{R}^d$ dont les composantes sont indépendantes centrées et réduites. La matrice de covariance de X est I_d. Le vecteur aléatoire AX est centré, de matrice de covariance $AI_dA^\top = \Sigma$. □

Le produit de Hadamard $A \circ B$ de deux matrices A et B de dimension $d \times d$ est la matrice de dimension $d \times d$ définie par $(A \circ B)_{i,j} = A_{i,j}B_{i,j}$ pour tout $1 \leq i, j \leq d$.

Corollaire 4.27 (Schur). *Si A et B sont deux matrices symétriques semi-définies positives de même dimension, alors $A \circ B$ est symétrique semi-définie positive.*

Démonstration. Soient X et Y deux vecteurs aléatoires indépendants et centrés de $\mathbb{R}^d$, de matrices de covariance respectives A et B. Le vecteur aléatoire Z de $\mathbb{R}^d$ défini par $Z_i = X_iY_i$ pour tout $1 \leq i \leq d$ est centré, et sa matrice de covariance est donnée par $A \circ B$ car X et Y sont indépendants et centrés. Ainsi, la matrice symétrique $A \circ B$ est semi-définie positive en tant que matrice de covariance d'un vecteur aléatoire ! □

4.3 Fonctions génératrices

Si (X, Y) est un vecteur aléatoire de $\mathbb{N}^2$, sa fonction génératrice $g_{(X,Y)} : [0, 1]^2 \to \mathbb{R}$ est

$$g_{(X,Y)}(s, t) = \mathbb{E}(s^X t^Y) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} s^n t^m \mathbb{P}(X = n, Y = m)$$

(en particulier, $g_{X+Y}(s) = g_{(X,Y)}(s, s)$). Le théorème qui suit montre que la fonction génératrice caractérise la loi du vecteur et permet de détecter l'indépendance des composants.

Théorème 4.28 (Fonctions génératrices). *Si (X, Y) est un vecteur aléatoire de $\mathbb{N}^2$ alors*

1. $\partial^n_{s=0}\partial^m_{t=0} g_{(X,Y)} = \mathbb{P}(X = n, Y = m)$ *pour tous* $n, m \in \mathbb{N}$
2. *X et Y sont indépendantes ssi* $g_{(X,Y)}(s, t) = g_X(s)g_Y(t)$ *pour tous* $(s, t) \in [0, 1]$
3. *si X et Y sont indépendantes alors* $g_{X+Y} = g_X g_Y$

Démonstration. Propriété 1. Immédiate. Propriété 2. Si X et Y sont indépendantes alors

$$g_{(X,Y)}(s, t) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X)\mathbb{E}(s^Y) = g_X(s)g_Y(t)$$

pour tout $s, t \in [0, 1[$, tandis que pour $s = 1$ ou $t = 1$ le résultat est immédiat. Réciproquement, si $g_{(X,Y)}(s, t) = g_X(s)g_Y(t)$ pour tous $s, t \in [0, 1[$ alors

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} s^n t^m \mathbb{P}(X = n, Y = m) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} s^n t^m \mathbb{P}(X = n)\mathbb{P}(Y = m)$$

ce qui donne $\mathbb{P}(X = n, Y = m) = \mathbb{P}(X = n)\mathbb{P}(Y = m)$ pour tous $n, m \in \mathbb{N}$ en prenant la dérivée $\partial^n_{s=0}\partial^m_{t=0}$ des deux membres (propriété 2.). Ainsi X et Y sont indépendantes.

Propriété 3. Découle de la propriété 2. avec $s = t$ car $g_{(X,Y)}(s, s) = g_{X+Y}(s)$. □

Exercice 4.29. *Soient X et Y deux v.a.r. sur $\mathbb{N}$ indépendantes. On a alors :*

1. *si* $X \sim \mathrm{Binom}(n, p)$ *et* $Y \sim \mathrm{Binom}(m, p)$ *alors* $X + Y \sim \mathrm{Binom}(n + m, p)$
2. *si* $X \sim \mathrm{Poi}(\lambda)$ *et* $Y \sim \mathrm{Poi}(\mu)$ *alors* $X + Y \sim \mathrm{Poi}(\lambda + \mu)$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Chapitre 5

Théorèmes limites

Ce chapitre concerne des suites de v.a.r. indépendantes.

5.1 Loi des grands nombres

Considérons un référendum dans une population de grande taille, qu'on modélise[1] par une suite de v.a.r. $X_1, X_2, \ldots$ indépendantes et de même loi[2] de Bernoulli de paramètre inconnu $p \in [0,1]$. On a donc $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = p = \mathbb{E}(X_n)$ pour tout $n \geq 1$. La loi faible des grands nombres affirme que la moyenne empirique $\frac{1}{n}(X_1 + \cdots + X_n)$ est un estimateur convergent et sans biais de p (convergence en probabilité).

Théorème 5.1 (Loi faible des grands nombres). *Soient $X_1, X_2, \ldots$ des v.a.r. indépendantes et de même loi, de carré intégrable. Si m est leur espérance, alors pour tout $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \varepsilon\right) = 0.$$

En d'autres termes, la suite $(X_n)_{n\geq 1}$ converge en probabilité vers m.

Si la loi des $X_1, X_2, \ldots$ est à support compact alors elles possèdent des moments de tout ordre, et en particulier ces v.a.r. sont de carré intégrable (ceci comprend le cas Bernoulli).

La loi faible doit sont nom à la fois à la nature de la convergence (en probabilité, alors que la loi forte énonce une convergence presque sûre), ainsi qu'à la nature des hypothèses (carré intégrable, alors que la loi forte ne nécessite que l'existence du moment d'ordre 1).

Démonstration. En utilisant l'inégalité de Bienaymé-Tchebychev (théorème 3.46) pour la v.a.r. $S_n = X_1 + \cdots + X_n - nm$, et le fait que $\sigma^2(S_n) = n\sigma^2(X_1)$ (hypothèses !), il vient

$$\begin{aligned}\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \varepsilon\right) &= \mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq n\varepsilon) \\ &\leq \frac{\sigma^2(S_n)}{n^2\varepsilon^2} = \frac{n\sigma^2(X_1)}{n^2}\varepsilon^2 = \mathcal{O}_{n\to\infty}\left(\frac{1}{n}\right).\end{aligned}$$

□

Exercice 5.2 (Contre exemple des lois de Cauchy). *Montrer que si $X_1, \ldots, X_n$ sont des v.a.r. indépendantes de loi de Cauchy alors leur moyenne empirique $\frac{1}{n}(X_1 + \cdots + X_n)$ suit*

1. Un sondage se modéliserait avec la loi hypergéométrique.
2. On dit aussi i.i.d. : indépendantes et identiquement distribuées.

Copyrighted material

également la loi de Cauchy. Indication : se ramener par récurrence au cas $n = 2$, puis calculer un produit de convolution en effectuant une décomposition en éléments simples. En déduire que la loi des grands nombres peut ne pas avoir lieu sans l'hypothèse d'intégrabilité. Cette propriété, mise en évidence sur les lois de Cauchy, concerne en réalité toute une classe de lois à queues lourdes (par exemple lois de Pareto ou Student).

Exemple 5.3 (Polynômes de Bernstein et théorème de Weierstrass). *Le théorème de Weierstrass sur la densité des polynômes $\mathbb{R}[X]$ dans $\mathcal{C}([a,b],\mathbb{R})$ pour la norme uniforme $\|\cdot\|_\infty$ peut être établi en utilisant la loi faible des grands nombres. En effet, un simple argument de translation et dilatation permet de se ramener au cas où $[a,b] = [0,1]$. À présent, fixons f dans $\mathcal{C}([0,1],\mathbb{R})$. Les polynômes de Bernstein $(P_n)_{n\geq 1}$ définis par*

$$P_n(X) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) X^k (1-X)^{n-k}$$

convergent uniformément vers f sur $[0,1]$. En effet, pour tout $x \in [0,1]$ et tout $n \in \mathbb{N}^*$,

$$P_n(x) = \mathbb{E}\left(f\left(\frac{S_n}{n}\right)\right) \quad \text{où} \quad S_n \sim \mathrm{Binom}(n,x).$$

Par conséquent, pour tout $x \in [0,1]$ et $n \in \mathbb{N}^*$,

$$f(x) - P_n(x) = \mathbb{E}\left(f(x) - f\left(\frac{S_n}{n}\right)\right).$$

Fixons un $\varepsilon > 0$ arbitrairement petit. Comme f est continue sur l'intervalle compact $[0,1]$, elle est uniformément continue d'après le théorème de Heine, et donc il existe $\eta > 0$ tel que $|f(x) - f(y)| \leq \varepsilon$ pour tous $x, y \in [0,1]$ tels que $|x - y| \leq \eta$. Fixons également $x \in [0,1]$ et considérons l'évènement $A_n = \left\{\left|\frac{S_n}{n} - x\right| \leq \eta\right\}$. La loi faible des grands nombres fournit un entier N tel que $\mathbb{P}(A_n^c) \leq \varepsilon$ pour tout $n \geq N$. Par conséquent, on a, pour tout $n \geq N$,

$$\begin{aligned}
|f(x) - P_n(x)| &\leq \mathbb{E}\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right) \\
&= \mathbb{E}\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \mathbf{1}_{A_n}\right) + \mathbb{E}\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \mathbf{1}_{A_n^c}\right) \\
&\leq \mathbb{E}(\varepsilon) + \mathbb{E}(2\|f\|_\infty \mathbf{1}_{A_n^c}) \\
&\leq (1 + 2\|f\|_\infty)\varepsilon
\end{aligned}$$

et cette borne est uniforme en $x \in [0,1]$, CQFD. Remarque : le théorème de Weierstrass permet d'établir, en utilisant la caractérisation de la loi par les fonctions tests continues et bornées, que si X et Y sont deux v.a.r. bornées avec une suite de moments identique, c'est-à-dire que $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ pour tout $n \geq 0$, alors X et Y ont même loi.

Théorème 5.4 (Loi forte des grands nombres - Hors programme). *Si $(X_n)_{n\geq 1}$ est une suite de v.a.r. indépendantes et de même loi possédant une espérance m alors*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = m\right) = 1.$$

En d'autres termes, la suite $(X_n)_{n\geq 1}$ converge presque sûrement vers m.

La preuve (hors programme) fait appel à une méthode de troncature pour se ramener à des variables bornées. Elle se trouve par exemple dans le livre de Paul S. Toulouse ou de

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Feller. Nous donnons ci-dessous deux preuves rapides lorsque les variables sont bornées (par une constante ou dans L^4) ce qui nous dispense de troncature. Il est également possible d'établir que pour une suite $X_1, X_2, \ldots$ de v.a.r. indépendantes et de même loi, la condition d'intégrabilité $\mathbb{E}(|X_1|) < \infty$ (i.e. X_1 possède une espérance) est nécessaire et suffisante pour que la loi forte des grands nombres ait lieu. Notons enfin que l'événement $\{\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = m\}$ est bien mesurable (exercice de réécriture dénombrable !).

Démonstration. Quitte à remplacer les X_i par $X_i - \mathbb{E}(X_i)$, on peut supposer que $m = 0$. On pose $S_n = X_1 + \cdots + X_n$. À présent, on observe qu'il suffit d'établir que

$$\mathbb{P}(\overline{\lim_n}\{|S_n| > n\varepsilon\}) = 0.$$

En effet, dans ce cas, pour une suite $\varepsilon_k \searrow 0$ arbitraire, on a $\mathbb{P}(\underline{\lim}_n\{|S_n \le n\varepsilon_k\}) = 0$ pour tout $k \ge 1$, et donc $\mathbb{P}(\lim_n n^{-1}S_n = 0) = \mathbb{P}(\cap_k \underline{\lim}_n\{|S_n| \le n\varepsilon_k\}) = 1$.

Preuve lorsque les variables sont bornées. On suppose qu'il existe une constante $C > 0$ telle que $\mathbb{P}(|X_i| \le C) = 1$ (ne dépend pas de i car les variables ont même loi). La première partie du lemme 2.31 de Borel-Cantelli permet de se ramener à établir que

$$\sum_{n\ge 1} \mathbb{P}(|S_n| \ge \varepsilon n) < \infty$$

pour tout $\varepsilon > 0$ fixé. Or pour tout $r > 0$ et tout entier $n \ge 1$, l'inégalité de Markov donne

$$\mathbb{P}(S_n \ge n\varepsilon) \le \mathbb{P}(e^{rS_n} \ge rn\varepsilon) \le e^{-rn\varepsilon}\mathbb{E}(e^{rS_n}) = e^{rn\varepsilon}\mathbb{E}(e^{rX_1})^n,$$

où l'égalité finale provient du fait que les variables $X_1, X_2, \ldots$ sont indépendantes et de même loi. À présent, comme $m = \mathbb{E}(X_1) = 0$ et $\mathbb{P}(|X_1| \le C) = 1$, il vient, en utilisant l'inégalité élémentaire $e^t - t \le e^{2t^2}$ si $t \in [0, 1/2]$ et $e^t - t \le e^t \le e^{2t^2}$ si $t \ge 1/2$,

$$\mathbb{E}(e^{rX_1}) \le e^{rC} - rC \le e^{2r^2C^2}.$$

Par conséquent, on obtient, en effectuant le choix optimal $r = \varepsilon/(4C^2)$,

$$\mathbb{P}(S_n \ge n\varepsilon) \le e^{rn\varepsilon - 2nr^2C^2} \le e^{-n\varepsilon^2/(8C^2)}.$$

Combinée à la même inégalité pour les variables $-X_1, -X_2, \ldots$, on obtient enfin

$$\mathbb{P}(|S_n| \ge n\varepsilon) \le 2e^{-n\varepsilon^2/(8C^2)}.$$

Le membre de droite est bien le terme général d'une série convergente, comme désiré.

Preuve lorsque les variables sont bornées dans L^4. On suppose que les variables $X_1, X_2, \ldots$ sont bornées dans L^4, c'est à dire que $\tau^4 = \mathbb{E}(X_1^4) < \infty$. On a alors

$$\mathbb{E}(S_n^4) = n\tau^4 + 3n(n-1)\sigma^4 = \mathcal{O}(n^2).$$

Ainsi, pour tout $\varepsilon > 0$, par l'inégalité de Markov,

$$\sum_n \mathbb{P}(|S_n| > n\varepsilon) = \sum_n \mathbb{P}(|S_n|^4 > n^4\varepsilon^4) \le \sum_n \frac{\mathbb{E}(S_n^4)}{n^4\varepsilon^4} < \infty$$

et le résultat découle à présent de la première partie du lemme 2.31 de Borel-Cantelli. Notons que si $\mathbb{P}(|X_1| \le C) = 1$ alors la suite $X_1, X_2, \ldots$ est bornée dans L^4, et la seconde preuve est donc plus puissante. D'autre part, elle reste valable même si les variables ne sont pas de même loi, pourvu qu'elles soient indépendantes et bornées dans L^4. □

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Exercice 5.5 (De la loi forte à la loi faible). *La loi faible (théorème 5.1) découle de la loi forte (théorème 5.4) car* $\mathbb{P}(n^{-1}|S_n| > \varepsilon) \leq \mathbb{P}(A_n)$ *avec* $A_n = \{\sup_{k\geq n} k^{-1}|S_k| > \varepsilon\}$ *et comme la suite* $(A_n)_{n\geq 1}$ *est croissante, on a* $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cap_n A_n) = 1$ *quand* $m = 0$.

Exemple 5.6 (Marche aléatoire simple sur $\mathbb{Z}$). *La marche aléatoire simple peut modéliser la position d'une particule dans un fluide, et sa version continue, le mouvement Brownien, a été introduite en physique par Einstein et Langevin, et en finance mathématique par Bachelier. Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.r. indépendantes et de même loi de Rademacher de paramètre* $p \in [0,1]$, *c'est-à-dire que* $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = -1) = p$ *pour tout* $n \geq 1$, *modélisant les incréments dus au choc avec les particules du fluide. La marche aléatoire simple sur* $\mathbb{Z}$ *est la suite* $(S_n)_{n\geq 1}$ *où* $S_n = X_1 + \cdots + X_n$. *La loi forte des grands nombres entraîne que si* $p \neq 1/2$ *alors* $(S_n)_{n\geq 1}$ *diverge vers* $\pm\infty$ *avec probabilité* 1 *quand* $n \to \infty$.

Remarque 5.7 (Une preuve de l'inégalité de Jensen avec la loi des grands nombres). *L'inégalité de Jensen affirme que si* X *est une v.a.r. intégrable et* $\varphi : \mathbb{R} \to \mathbb{R}$ *une fonction convexe telle que* $\varphi(X)$ *est intégrable, alors*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

En particulier, pour $\varphi(x) = x^2$ *on retrouve* $\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)$ *tandis que pour* $\varphi(x) = |x|$ *on retrouve* $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$. *Pour établir l'inégalité de Jensen, on commence par exprimer la convexité de* φ *: pour tout entier* $n \geq 1$ *et tous réels* $x_1, \ldots, x_n$,

$$\varphi\left(\frac{x_1 + \cdots + x_n}{n}\right) \leq \frac{\varphi(x_1) + \cdots + \varphi(x_n)}{n}.$$

Nous pouvons appliquer cette inégalité en remplaçant $x_1, \ldots, x_n$ *par une suite* $X_1, \ldots, X_n$ *de v.a.r. indépendantes de même loi que* X *(l'inégalité est valable pour tout* $\omega \in \Omega$*). À présent, la loi forte des grands nombres appliquée deux fois affirme qu'avec probabilité* 1,

$$\lim_{n\to\infty} \varphi\left(\frac{X_1 + \cdots + X_n}{n}\right) = \varphi(\mathbb{E}(X)) \quad \text{et} \quad \lim_{n\to\infty} \frac{\varphi(X_1) + \cdots + \varphi(X_n)}{n} = \mathbb{E}(\varphi(X)).$$

La première convergence fait appel à la continuité de φ, *qui découle de sa convexité. Enfin, si* A *et* B *sont deux événements tels que* $\mathbb{P}(A) = \mathbb{P}(B) = 1$ *alors* $A \cap B \neq \varnothing$.

Remarque 5.8 (Convergence monotone ou théorème de Fubini-Tonelli). *Le lecteur familier avec l'intégrale de Lebesgue connaît bien le théorème de convergence monotone : si* $(X_n)_{n\geq 1}$ *est une suite croissante de v.a.r. à valeurs dans* $[0,\infty]$ *alors*

$$\lim_n \mathbb{E}(X_n) = \mathbb{E}(\lim_n X_n).$$

D'autre part, si X *est une v.a.r. sur* $[0,\infty]$ *vérifiant* $\mathbb{E}(X) < \infty$ *alors* $\mathbb{P}(X < \infty) = 1$. *La première partie du lemme de Borel-Cantelli en découle car*

$$\sum_n \mathbb{P}(A_n) = \sum_n \mathbb{E}(\mathbf{1}_{A_n}) = \mathbb{E}\left(\sum_n \mathbf{1}_{A_n}\right) = \mathbb{E}\left(\mathbf{1}_{\overline{\lim}_n A}\right).$$

On peut également voir ce résultat comme une conséquence du théorème de Fubini-Tonelli plutôt que comme une application du théorème de convergence monotone. Une autre conséquence du théorème de convergence monotone (ou du théorème de Fubini-Tonelli) est que

$$\sum_n \mathbb{E}(|Y_n|) < \infty \quad \Rightarrow \quad \mathbb{P}(\lim_n Y_n = 0) = 1.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

En effet, on a $\mathbb{E}(\sum_n |Y_n|) = \sum_n \mathbb{E}(|Y_n|) < \infty$ *et donc* $\sum_n |Y_n|$ *est une v.a.r. sur* $[0, \infty]$ *d'espérance finie, et donc finie avec probabilité* 1*, ce qui implique que* $|Y_n|$ *tend vers* 0 *avec probabilité* 1*. Cette observation suggère une preuve alternative du théorème 5.4 :*

$$\mathbb{E}\left(\sum_n \left(\frac{S_n}{n}\right)^4\right) = \sum_n \mathbb{E}\left(\left(\frac{S_n}{n}\right)^4\right) < \infty \quad \textit{d'où} \quad \mathbb{P}\left(\lim_n \frac{S_n}{n} = 0\right) = 1.$$

Remarque 5.9 (Suite). *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.r. indépendantes de carré intégrable. Si* $\sum_n \mathrm{Var}(X_n) < \infty$ *alors en vertu du théorème de convergence monotone ou du théorème de Fubini-Tonelli positif on a* $\mathbb{P}(\lim_n X_n - \mathbb{E}(X_n) = 0) = 1$*. En particulier, si* $(X_n)_{n\geq 1}$ *sont des v.a.r. i.i.d. centrées de carré intégrable alors*

$$\mathbb{P}(\lim_n n^{-1} X_n = 0) = 1.$$

Attention, il ne s'agit pas de la loi des grands nombres, qui concerne la quantité $n^{-1}(X_1 + \cdots + X_n)$*.*

Remarque 5.10 (Statistique d'ordre et vecteur des rang). *Soient* $X_1, \ldots, X_n$ *des v.a.r. indépendantes et de même loi admettant une densité. L'hypothèse de densité entraîne qu'avec probabilité* 1 *les composantes du vecteur* $X = (X_1, \ldots, X_n)$ *sont deux à deux différentes, et qu'il existe donc une unique permutation (aléatoire)* π_X *à valeur sur le groupe symétrique* $\mathcal{S}_n$ *telle que* $X_{\pi_X(1)} < \cdots < X_{\pi_X(n)}$*. On dit qu'il s'agit de la* statistique d'ordre de l'échantillon $X_1, \ldots, X_n$*. En particulier*

$$X_{\pi_X(1)} = \min(X_1, \ldots, X_n) \quad \textit{et} \quad X_{\pi_X(n)} = \max(X_1, \ldots, X_n).$$

On dit que le vecteur $(\pi_X^{-1}(1), \ldots, \pi_X^{-1}(n))$ *est le* vecteur des rangs de l'échantillon $X_1, \ldots, X_n$*. La loi du vecteur* X *est échangeable, c'est-à-dire que les vecteurs aléatoires* X *et* $X_\sigma = (X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ *ont la même loi quelque soit* $\sigma \in \mathcal{S}_n$*, d'où*

$$\mathbb{P}(\pi_X = \sigma) = \mathbb{P}(X_{\sigma(1)} < \cdots < X_{\sigma(n)}) = \mathbb{P}(X_1 < \cdots < X_n).$$

Le membre de droite ne dépend pas du choix de $\sigma \in \mathcal{S}_n$ *et vaut donc* $1/\mathrm{card}(S_n) = 1/n!$*. Ainsi,* σ *suit la loi uniforme sur* $\mathcal{S}_n$*. Alternativement, il est également possible d'observer que* $\pi_{X_\sigma} = \pi_X \circ \sigma$ *quelque soit* $\sigma \in \mathcal{S}_n$*, donc que la loi de* π_X *est invariante par toute translation, et donc que* π_X *suit la loi uniforme sur* $\mathcal{S}_n$ *en vertu de l'exemple 2.13. Il se trouve que* π_X *et la statistique d'ordre* X_{π_X} *sont indépendantes, car pour tout borélien* $A \subset \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 < \cdots < x_n\}$ *et toute permutation* $\sigma \in \mathcal{S}_n$*,*

$$\mathbb{P}(\pi_X = \sigma, X_{\pi_X} \in A) = \mathbb{P}(X_{\sigma(1)} < \cdots < X_{\sigma(n)}, X_\sigma \in A) = \frac{1}{n!}\mathbb{P}(X \in \{A_{\sigma'} : \sigma' \in \mathcal{S}_n\}).$$

Exemple 5.11 (Covariance empirique). *Soit* X *un vecteur colonne aléatoire de* $\mathbb{R}^m$ *centré et de matrice de covariance* Σ*. Soient* $X_1, \ldots, X_n$ *des vecteurs colonne aléatoires de* $\mathbb{R}^m$ *de même loi que* X*. La* matrice de covariance empirique Σ_n *est définie par*

$$\Sigma_n = \frac{1}{n}(X_1 X_1^\top + \cdots + X_n X_n^\top) = \frac{1}{n}\mathbb{X}^\top \mathbb{X}$$

où $\mathbb{X}$ *est la matrice aléatoire* $m \times n$ *dont les lignes sont* $X_1^\top, \ldots, X_n^\top$*. La matrice aléatoire* Σ_n *est symétrique semi-définie positive car combinaison convexe de telles matrices. On a* $\mathbb{E}(\Sigma_n) = \Sigma$ *et la loi forte des grands nombres indique que* Σ_n *converge entrée par entrée vers* Σ*, avec probabilité* 1*. Comme l'ensemble des matrices inversible est ouvert, si* Σ *est inversible, alors avec probabilité* 1*, la matrice* Σ_n *est inversible pour* n *assez grand.*

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

5.2 Théorème de la limite centrale

La loi forte des grands nombres montre que la moyenne empirique

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$$

est un estimateur consistant et sans biais de m. Pour tout $0 \leq \alpha \leq 1/2$, le théorème 5.12 de la limite centrale fournit un intervalle de confiance pour m de niveau asymptotique α :

$$I_{n,\alpha} = \left[\overline{X}_n - \frac{\sigma q_\alpha}{\sqrt{n}}, \overline{X}_n + \frac{\sigma q_\alpha}{\sqrt{n}}\right]$$

où q_α est le quantile $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite. On a en effet

$$\mathbb{P}(m \in I_{n,\alpha}) = \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \in [-q_\alpha, q_\alpha]\right) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-q_\alpha}^{q_\alpha} e^{-\frac{u^2}{2}}\, du = 1 - \alpha.$$

L'intervalle $I_{n,\alpha}$ est de largeur $2\sigma q_\alpha n^{-1/2}$, qui tend vers 0 si $n \to \infty$, et vers ∞ si $\alpha \to 0$.

Théorème 5.12 (Théorème de la limite centrale – Admis). *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi, de variance non nulle et finie σ^2 et de moyenne m. Alors pour tout intervalle I de $\mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \cdots + X_n}{n} - m\right) \in I\right) = \frac{1}{\sqrt{2\pi}} \int_I e^{-\frac{u^2}{2}}\, du.$$

En d'autre termes, en posant $S_n = X_1 + \cdots + X_n$, la variable aléatoire

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - m\right) = \frac{S_n - nm}{\sqrt{n\sigma^2}} = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\sigma^2(S_n)}}$$

converge en loi vers la loi normale centrée réduite $\mathcal{N}(0,1)$ lorsque $n \to \infty$.

L'expression $(S_n - \mathbb{E}(S_n))/\sigma(S_n)$ montre bien que la moyenne est fixée à 0 et l'écart-type à 1 le long du théorème de la limite centrale. La vitesse en $\sqrt{n}$ provient de $\sigma(S_n) = \sigma\sqrt{n}$. Si les (X_n) sont de loi normale $\mathcal{N}(m, \sigma^2)$ alors $(S_n - nm)/\sqrt{n\sigma^2} \sim \mathcal{N}(0,1)$ et cela permet de concevoir le théorème de la limite centrale comme une sorte de théorème du point fixe.

Démonstration. La preuve basée sur les fonctions caractéristiques est rapide mais hors programme. Utilisons plutôt l'inégalité de couplage de Lindeberg du lemme 5.13 dans le cas où $\mathbb{E}(|X_1 - m|^3) < \infty$, ce qui couvre le cas Bernoulli. Pour toute fonction $f \in \mathcal{C}^3(\mathbb{R}, \mathbb{R})$ telle que f, f', f'', f''' sont bornées, et tout $n \geq 1$, ce lemme donne

$$\left|\mathbb{E}\left(f\left(\frac{S_n - nm}{\sqrt{n}\sigma}\right)\right) - \int_{\mathbb{R}} f(u) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du\right| \leq \frac{\|f'''\|_\infty \mathbb{E}(|X_1 - m|^3)}{2\sqrt{n}\sigma^3}.$$

Pour tout $a \in \mathbb{R}$, on pose $I_a =]-\infty, a]$. Pour tout $\varepsilon > 0$, on peut construire $f_{a,\varepsilon} \in \mathcal{C}^3(\mathbb{R}, \mathbb{R})$ avec $f_{a,\varepsilon}, f'_{a,\varepsilon}, f''_{a,\varepsilon}, f'''_{a,\varepsilon}$ bornées, telle que $\mathbf{1}_{I_{a+\varepsilon}} \leq f_{a,\varepsilon} \leq \mathbf{1}_{I_a}$ et $\left\|f'''_{a,\varepsilon}\right\|_\infty \leq \varepsilon^{-3}$. De plus,

$$\left|\int_{I_a} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du - \int_{I_b} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du\right| \leq \frac{1}{\sqrt{2\pi}} |b - a| \leq |b - a|.$$

Cela donne, pour tout $c \in \mathbb{R}$, tout $n \geq 1$ et tout $\varepsilon > 0$, en utilisant $f_{c\pm\varepsilon,\varepsilon}$,

$$-\frac{\mathbb{E}(|X_1 - m|^3)}{2\sqrt{n}\sigma^3\varepsilon^3} - \varepsilon \leq \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \in I_c\right) - \int_{I_c} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \leq \frac{\mathbb{E}(|X_1 - m|^3)}{2\sqrt{n}\sigma^3\varepsilon^3} + \varepsilon.$$

□

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Lemme 5.13 (Inégalité de couplage de Lindeberg). *Soient $X_1, Y_1, \dots, X_n, Y_n$ des v.a.r. indépendantes telles que $\mathbb{E}(|X_k|^3) < \infty$ et $Y_k \sim \mathcal{N}(\mathbb{E}(X_k), \sigma^2(X_k))$ pour tout $1 \le k \le n$. Alors pour toute $f \in \mathcal{C}^3(\mathbb{R}, \mathbb{R})$ avec f, f', f'', f''' bornées, en posant $\tau_k^3 = \mathbb{E}(|X_k - \mathbb{E}(X_k)|^3)$,*

$$|\mathbb{E}(f(X_1 + \cdots + X_n)) - \mathbb{E}(f(Y_1 + \cdots + Y_n))| \le \frac{(\tau_1^3 + \cdots + \tau_n^3)}{2} \|f'''\|_\infty.$$

Démonstration. Quitte à translater f on peut se placer dans le cas où $\mathbb{E}(X_k) = 0$ pour tout $1 \le k \le n$. Fixons $n \ge 1$ et posons $Z_k = X_1 + \cdots + X_{k-1} + Y_{k+1} + \cdots + Y_n$ pour tout $1 \le k \le n$, ce qui donne la somme téléscopique (penser à des dominos !)

$$\mathbb{E}(f(X_1 + \cdots + X_n)) - \mathbb{E}(f(Y_1 + \cdots + Y_n)) = \sum_{k=1}^{n} \mathbb{E}(f(Z_k + X_k) - f(Z_k + Y_k)).$$

La formule de Taylor-Lagrange appliquée à f à l'ordre 2 en Z_k donne

$$\left| f(Z_k + X_k) - f(Z_k) - f'(Z_k)X_k - f''(Z_k)\frac{X_k^2}{2!} \right| \le \frac{|X_k|^3}{3!} \|f'''\|_\infty$$

et

$$\left| f(Z_k + Y_k) - f(Z_k) - f'(Z_k)Y_k - f''(Z_k)\frac{Y_k^2}{2!} \right| \le \frac{|Y_k|^3}{3!} \|f'''\|_\infty.$$

Comme X_k et Y_k sont indépendantes de Z_k et ont des moments d'ordre 1 et 2 identiques,

$$|\mathbb{E}(f(Z_k + X_k) - f(Z_k + Y_k))| \le \frac{\|f'''\|_\infty}{3!} \mathbb{E}(|X_k|^3 + |Y_k|^3).$$

Comme $Y_k = \sigma(X_k)G_k$ avec $G_k \sim \mathcal{N}(0,1)$ et comme $\mathbb{E}(|G_k|^3) = 4/\sqrt{2\pi} \le 2$, on obtient

$$\mathbb{E}(|Y_k|^3) = \mathbb{E}(|X_k|^2)^{3/2}\mathbb{E}(|G_k|^3) \le 2\mathbb{E}(|X_k|^3).$$

□

Exemple 5.14 (Ruine du joueur). *Un joueur gagne à chaque partie 1 Euro avec probabilité p et perd 1 Euro avec probabilité $1-p$. Les parties sont indépendantes. Sa fortune à l'instant n est une v.a.r. X_n qui vérifie $X_0 = 0$ (fortune initiale nulle) et $X_n = \varepsilon_1 + \cdots + \varepsilon_n$ où $(\varepsilon_k)_{k\ge1}$ est une suite de v.a.r. indépendantes de même loi de Rademacher de paramètre p, c'est-à-dire que $\mathbb{P}(\varepsilon_k = 1) = 1 - \mathbb{P}(\varepsilon_k = -1) = p$ pour tout $k \ge 1$. On pose $m = 2p - 1$ et $\sigma^2 = 4p(1-p)$. Soit $a, b \in \mathbb{Z}$ avec $a < 0 < b$. Le joueur quitte le jeu dès qu'il est assez riche ou trop endetté, c'est-à-dire dès que sa fortune atteint b ou sa dette atteint a. Cela se produit à l'instant aléatoire suivant :*

$$T = \inf\{n \ge 1 : X_n \notin [a, b]\} = \inf\{n \ge 1 : X_n = a \text{ ou } X_n = b\}.$$

Montrons que $\mathbb{P}(T < \infty) = 1$. Si $p \ne 1/2$ alors $m \ne 0$ et par la loi forte des grands nombres, presque sûrement $(X_n)_{n\ge1}$ tend vers $+\infty$ si $p > 1/2$ et vers $-\infty$ si $p < 1/2$. Reste le cas $p = 1/2$ c'est-à-dire $m = 0$. Pour tout $n \ge 1$, en posant $I_n = \frac{1}{\sqrt{n}}[a, b[$, on a

$$\mathbb{P}(T = \infty) \le \mathbb{P}(a < X_n < b) = \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right).$$

Or $(n^{-1/2}X_n)_{n\ge1}$ converge en loi vers $\mathcal{N}(0, \sigma^2)$ par le théorème de la limite centrale. Mais I_n dépend de n. Cependant, comme $(I_n)_{n\ge1}$ est décroissante, on a

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right) \le \inf_{m\ge1} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{I_m} e^{-\frac{t^2}{2\sigma^2}}\, dt = 0.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Exercice 5.15 (Du théorème de la limite centrale à la loi faible des grands nombres). *Il est possible de déduire la loi faible des grands nombres du théorème de la limite centrale. En effet, si $(Y_n)_{n\geq 0}$ est une suite de v.a.r. qui converge en loi vers une loi $\mathcal{L}$ et si $(a_n)_{n\geq 0}$ est une suite déterministe qui tend vers 0, alors $(a_nY_n)_{n\geq 0}$ converge en probabilité vers 0. Pour le voir, en utilisant la caractérisation de la convergence en loi au moyen des fonctions test continues et bornées, établir que si une suite de v.a.r. converge en loi vers la loi d'une v.a.r. constante alors la convergence a lieu en probabilité, puis adapter la preuve, ce qui donne :*

$$\mathbb{P}(|a_nY_n - 0| \geq \varepsilon) \overset{n\gg 1}{\leq} \mathbb{P}(|Y_n| \geq 1) = \mathbb{E}(\mathbf{1}_{|-1,1|^c}(Y_n)) \leq \mathbb{E}(|Y_n|) \underset{n\to\infty}{\longrightarrow} \mathbb{E}(|0|) = 0.$$

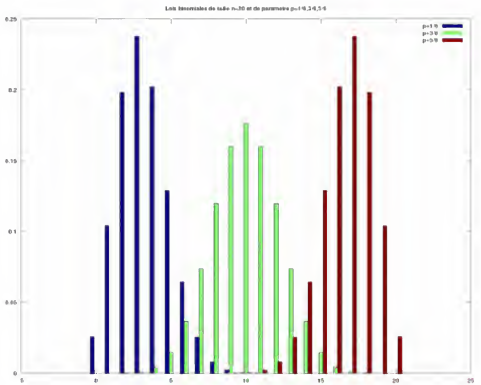


Figure 5.1 – La loi binomiale de taille n grand et de paramètre p est bien approchée par une loi normale lorsque p est proche de $1/2$ et par une loi de Poisson lorsque p et proche de 0 (lorsque p est proche de 1, il suffit de « renverser horizontalement la loi de Poisson ».

5.3 Approximation de la loi Binomiale par la loi normale

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi, de moyenne m et de variance non nulle et finie σ^2. On pose $S_n = X_1 + \cdots + X_n$ pour tout $n \geq 1$. Le théorème 5.12 de la limite centrale indique que pour tout $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \leq t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

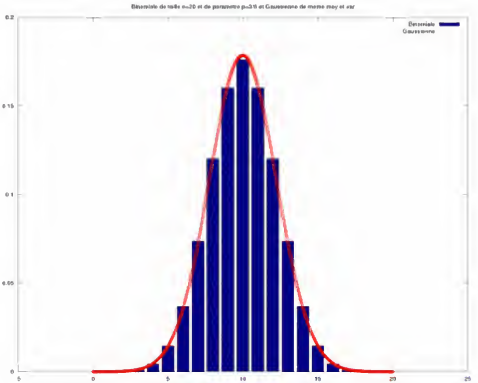


Figure 5.2 – Approximation de la loi binomiale par la loi normale.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Cela suggère que la loi de S_n est proche de $\mathcal{N}(nm, n\sigma^2)$ lorsque n est grand. Le théorème de Berry-Esseen précise cette proximité en affirmant que pour tout $t \in \mathbb{R}$ et $n \geq 1$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \leq t \right) - \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}\, dx \right| \leq \frac{\tau^3}{\sqrt{n}\sigma^3}$$

où $\tau^3 = \mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)$. Lorsque $(X_n)_{n\geq 1}$ sont de Bernoulli de paramètre $p \in]0,1[$, en notant $q = 1-p$ on trouve $m = p$, $\sigma^2 = pq$ et $\tau^3 = pq(1-2pq)$ ce qui donne

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}(S_n \leq \sqrt{npq}t + np) - \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}\, dx \right| \leq \frac{1-2pq}{\sqrt{npq}}.$$

Cette approximation de la loi binomiale par la loi normale est d'autant plus bonne que $(1-2p(1-p))/\sqrt{np(1-p)}$ est petit. À n fixé, cette borne est minimale pour $p = 1/2$ mais explose quand p se rapproche de 0 ou de 1. Une preuve du théorème de Berry-Esseen se trouve dans le livre de Feller. Notons que notre preuve du théorème 5.12 de la limite centrale permet déjà d'établir une borne en $\mathcal{O}(n^{-1/8})$ en y prenant $\varepsilon^4 = \mathcal{O}(n^{-1/2})$.

Il est possible de quantifier la proximité de la loi binomiale à la loi normale en utilisant la densité plutôt que la fonction de répartition, comme dans le théorème suivant.

Théorème 5.16 (de Moivre et Laplace). *Si $S_n \sim \mathrm{Binom}(n,p)$ avec $0 < p < 1$ et $q = 1-p$ alors pour tous $-\infty < a < b < +\infty$ on a la convergence uniforme suivante :*

$$\lim_{n\to\infty} \sqrt{n} \sup_{k\in I_n(a,b)} \left| \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq} \right) \right| = 0$$

où

$$I_n(a,b) = \left\{ 0 \leq k \leq n : \frac{k-np}{\sqrt{npq}} \in [a,b] \right\}.$$

Le théorème 5.16 fournit par intégration le TLC dans le cas Bernoulli :

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} \in [a,b] \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{s^2}{2}}\, dx.$$

Démonstration. Nous allons utiliser le raffinement suivant de la formule de Stirling :

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\theta_n} \quad \text{avec} \quad |\theta_n| \leq \frac{c}{n}$$

pour une constante $c > 0$. En posant $q = 1-p$ et $j = n-k$ il vient

$$\mathbb{P}(S_n = k) = p^k (1-p)^{n-k} \binom{n}{k} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{kj}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{j}\right)^j e^{\theta_n - \theta_k - \theta_j}.$$

On remarque tout d'abord que

$$|\theta_n - \theta_k - \theta_j| \leq c\left|\frac{1}{n} + \frac{1}{k} + \frac{1}{j}\right| = c\left(\frac{1}{n} + \frac{n}{kj}\right).$$

Fixons à présent $-\infty < a < b < +\infty$ et $x \in [a,b]$ de sorte que $k = np + x\sqrt{npq} \in I_n(a,b)$ et $x = (k-np)/\sqrt{npq} \in [a,b]$. On a alors $j = n-k = nq - x\sqrt{npq}$. On observe que

$$\lim_{n\to\infty} \sup_{x\in[a,b]} \left| \frac{1}{npq}\frac{kj}{n} - 1 \right| = \lim_{n\to\infty} \sup_{x\in[a,b]} \left| \left(\sqrt{\frac{p}{q}} + x\frac{1}{\sqrt{n}}\right)\left(\sqrt{\frac{q}{p}} - x\frac{1}{\sqrt{n}}\right) - 1 \right| = 0$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

En particulier, pour n assez grand,

$$\sup_{x\in[a,b]} |\theta_n - \theta_k - \theta_j| \leq c\left(\frac{1}{n} + \frac{2}{pqn}\right).$$

D'autre part, on a

$$k\ln\left(\frac{np}{k}\right) = (np + x\sqrt{npq})\ln\left(1 - \frac{x\sqrt{npq}}{np + x\sqrt{npq}}\right).$$

Comme[3] $\left|\ln(1+u) - u + \frac{1}{2}u^2\right| \leq \frac{|u|^3}{3(1-|u|)}$ pour tout $u > 1$ il vient

$$\lim_{n\to\infty} \sup_{x\in[a,b]} \left|k\ln\left(\frac{np}{k}\right) + x\sqrt{npq} + \frac{x^2npq}{2(np + x\sqrt{npq})}\right| = 0.$$

Comme

$$\frac{x^2npq}{2(np + x\sqrt{npq})} = \frac{x^2q}{2}\frac{1}{\left(1 + x\sqrt{\frac{q}{np}}\right)}$$

on a

$$\lim_{n\to\infty} \sup_{x\in[a,b]} \left|k\ln\left(\frac{np}{k}\right) + x\sqrt{npq} + \frac{x^2q}{2}\right| = 0.$$

De même on montre que

$$\lim_{n\to\infty} \sup_{x\in[a,b]} \left|j\ln\left(\frac{nq}{j}\right) - x\sqrt{npq} + \frac{x^2p}{2}\right| = 0.$$

On en déduit que

$$\mathbb{P}(S_n = k) = \frac{r_n(x)}{\sqrt{2\pi npq}}\exp\left(-\frac{x^2}{2}\right) \quad \text{où} \quad \lim_{n\to\infty} \sup_{x\in[a,b]} |r_n(x) - 1| = 0.$$

□

5.4 Distance en variation

Dans toute cette section, E est un ensemble au plus dénombrable. L'ensemble des lois sur E est un espace métrique pour la distance en variation (on dit aussi *variation totale*)

$$d_V(\mu,\nu) = \sup_{A\subset E} |\mu(A) - \nu(A)|.$$

On a $0 \leq d_V(\mu,\nu) \leq 1$ et $d_V(\mu,\nu) = 1$ si μ et ν ont des supports disjoints. Le théorème suivant affirme en particulier que la distance en variation coïncide avec la distance $\ell^1(E)$.

Théorème 5.17 (Expressions alternatives). *Si μ et ν sont des lois sur E alors*

$$d_V(\mu,\nu) = \frac{1}{2}\sup_{f:E\to[-1,1]} \left|\int f\,d\mu - \int f\,d\nu\right| = \frac{1}{2}\sum_{x\in E} |\mu(x) - \nu(x)|$$

De plus, le supremum dans la définition de d_V est atteint pour l'ensemble

$$A_* = \{x \in E : \mu(x) \geq \nu(x)\}$$

tandis que dans l'expression variationnelle fonctionnelle de d_V il est atteint pour

$$f = \mathbf{1}_{A_*} - \mathbf{1}_{A_*^c}.$$

3. Car $\ln(1+u) = u - \frac{1}{2}u^2 + \int_0^u t^2(1+t)^{-1}\,dt$ pour tout $u > -1$.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. La seconde égalité provient de l'identité

$$\left|\int f\,d\mu - \int f\,d\nu\right| \leq \sum_{x\in E} |f(x)||\mu(x)-\nu(x)| \leq \sup_{x\in E}|f(x)| \sum_{x\in E}|\mu(x)-\nu(x)|$$

qui est saturée pour $f = \mathbf{1}_{A_*} - \mathbf{1}_{A_*^c}$. Pour la première égalité, on écrit

$$|\mu(A)-\nu(A)| = \frac{1}{2}\left|\int f_A\,d\mu - \int f_A\,d\nu\right|$$

où $f = \mathbf{1}_A - \mathbf{1}_{A^c}$, ce qui donne

$$|\mu(A)-\nu(A)| \leq \frac{1}{2}\sup_{f:E\to[-1,1]}\left|\int f\,d\mu - \int f\,d\nu\right| = \frac{1}{2}\sum_{x\in E}|\mu(x)-\nu(x)|$$

qui est saturé pour $A = A_*$ car

$$2|\mu(A_*)-\nu(A_*)| = \sum_{x\in A_*}|\mu(x)-\nu(x)| + \sum_{x\in A_*^c}|\mu(x)-\nu(x)|.$$

□

Théorème 5.18 (Convergence en loi). *Si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires sur E et si μ_n désigne la loi de X_n, alors pour toute loi μ sur E, les propriétés suivantes sont équivalentes :*

1. $\lim_{n\to\infty}\int f\,d\mu_n = \int f\,d\mu$ *pour toute fonction bornée* $f : E \to \mathbb{R}$
2. $\lim_{n\to\infty}\mu_n(x) = \mu(x)$ *pour tout* $x \in E$
3. $\lim_{n\to\infty} d_V(\mu_n,\mu) = 0$

Lorsqu'elles ont lieu on dit que (X_n) converge en loi vers μ quand $n \to \infty$.

Démonstration. Pour déduire *1.* de *3.* il suffit d'utiliser l'expression variationnelle fonctionnelle de d_V. Pour déduire *2.* de *1.* on peut prendre $f = 1_{\{x\}}$. Pour déduire *3.* de *2.* on observe que pour tout $A \subset E$,

$$\sum_{x\in E}|\mu_n(x)-\mu(x)| = \sum_{x\in A}|\mu_n(x)-\mu(x)| + \sum_{x\in A^c}|\mu_n(x)-\mu(x)|,$$

ensuite, grâce à *4.*, pour tout $\varepsilon' > 0$ il existe un entier $N = N(A,\varepsilon')$ tel que le premier terme du membre de droite est majoré par $\mathrm{card}(A)\varepsilon'$ pour tout $n \geq N$. Pour le second terme du membre de droite, on écrit

$$\sum_{x\in A^c}|\mu_n(x)-\mu(x)| \leq \sum_{x\in A^c}\mu_n(x) + \sum_{x\in A^c}\mu(x).$$

Puisqu'on a

$$\sum_{x\in A^c}\mu_n(x) = \sum_{x\in A}\mu(x) - \sum_{x\in A}\mu_n(x) + \sum_{x\in A^c}\mu(x)$$

on obtient

$$\sum_{x\in A^c}|\mu_n(x)-\mu(x)| \leq \sum_{x\in A}|\mu_n(x)-\mu(x)| + 2\sum_{x\in A^c}\mu(x).$$

Puisque $\mu \in \mathcal{P}$, pour tout $\varepsilon'' > 0$, on peut choisir A fini tel que $\mu(A^c) \leq \varepsilon''$. □

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Remarque 5.19 (Dispersion à l'∞). *Si (μ_n) sont des lois et $\mu(x) = \lim_{n\to\infty} \mu_n(x)$ alors μ n'est pas forcément une loi, sauf si E est fini. En effet, lorsque E est infini, il peut se produire un phénomène de dispersion de la masse à l'infini. Contre exemple : $E = \mathbb{N}$ et μ_n affecte la masse $1/n$ aux singletons $\{1\}, \ldots, \{n\}$, ce qui donne μ identiquement nulle.*

Théorème 5.20 (Autre expression et cas extrémal). *Si μ et ν sont des lois sur E alors*

$$d_V(\mu, \nu) = 1 - \sum_{x\in E} (\mu(x) \wedge \nu(x)).$$

En particulier, $d_V(\mu, \nu) = 1$ si et seulement si μ et ν ont des support disjonits.

Démonstration. Il suffit d'écrire

$$\sum_{x\in E} (\mu(x) \wedge \nu(x)) = \frac{1}{2} \sum_{x\in E} (\mu(x) + \nu(x) - |\mu(x) - \nu(x)|) = 1 - d_V(\mu, \nu).$$

□

Théorème 5.21 (Couplage). *Si μ et ν sont des lois sur E alors*

$$d_V(\mu, \nu) = \inf_{(X,Y)} \mathbb{P}(X \neq Y)$$

où l'infimim porte sur les couples de v.a. sur $E \times E$ de lois marginales μ et ν. De plus, il existe un couple de ce type pour lequel l'égalité est atteinte (i.e. l'infimum est un minimum).

Démonstration. Soit (X, Y) est un couple de variables aléatoires sur $E\times E$ de lois marginales μ et ν. Comme $\mathbb{P}(X = x, Y = x) \leq \mu(x) \wedge \nu(x)$ pour tout $x \in E$ on a

$$1 - d_V(\mu, \nu) = \sum_{x\in E} (\mu(x) \wedge \nu(x)) \geq \sum_{x\in E} \mathbb{P}(X = x, Y = x) = \mathbb{P}(X = Y).$$

Il suffit donc de construire un couple (X, Y) pour lequel l'égalité est atteinte. Posons

$$p = 1 - d_V(\mu, \nu) \in [0, 1].$$

Cas où $p = 0$. On a alors $d_V(\mu, \nu) = 1$ et μ et ν ont des supports disjoints. Cela donne $\mathbb{P}(X = Y) = \sum_{x\in E} \mu(x)\nu(x) = 0$. On prend (X, Y) avec $X \sim \mu$ et $Y \sim \nu$ indépendantes.
Cas où $p = 1$. On a alors $d_V(\mu, \nu) = 0$ et donc $\mu = \nu$. On prend (X, X) où $X \sim \mu$.
Cas où $0 < p < 1$. Soit (U, V, W) un triplet de variables aléatoire de lois respectives

$$p^{-1}(\mu \wedge \nu), \quad (1-p)^{-1}(\mu - (\mu \wedge \nu)), \quad (1-p)^{-1}(\nu - (\mu \wedge \nu)).$$

Notons que $p = \sum_{x\in E}(\mu(x) \wedge \nu(x))$. Soit B une variable aléatoire de Bernouill, indépendante de (U, V, W), telle que $\mathbb{P}(B = 1) = 1 - \mathbb{P}(B = 0) = p$. Définissons $(X, Y) = (U, U)$ si $B = 1$ et $(X, Y) = (V, W)$ si $B = 0$. On a alors $X \sim \mu$ et $Y \sim \nu$, et puisque les lois de V et W ont des supports disjoints, on a $\mathbb{P}(V = W) = 0$, et donc $\mathbb{P}(X = Y) = \mathbb{P}(B = 1) = p$. □

5.5 Approximation de la loi Binomiale par la loi de Poisson

Si S_n suit la loi binomiale $\mathrm{Binom}(n, p)$ alors pour tout $k \in \mathbb{N}$, on a,

$$\mathbb{P}(S_n = k) - e^{-np}\frac{(np)^k}{k!} = \left(\frac{n}{n(1-p)} \cdots \frac{(n-k+1)}{n(1-p)}(1-p)^n - e^{-np}\right)\frac{(np)^k}{k!}.$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

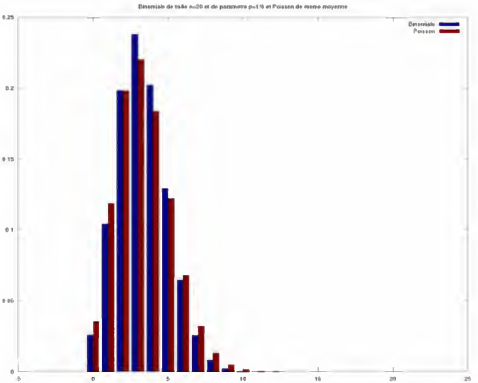


Figure 5.3 – Approximation de la loi binomiale par la loi de Poisson.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Ceci montre que si p dépend de n avec $\lim_{n\to\infty} np = \lambda$ alors la loi de S_n tend vers $\mathrm{Poi}(\lambda)$. La distance en variation permet de quantifier cette convergence en loi : l'inégalité de poissonisation de Le Cam du théorème 5.22 ci-dessous donne (utile si np^2 est petit)

$$\sum_{k=0}^{\infty} \left| \mathbb{P}(S_n = k) - e^{-np}\frac{(np)^k}{k!} \right| \leq 2np^2.$$

Théorème 5.22 (Inégalité de poissonisation de Le Cam). *Soient $X_1, \ldots, X_n$ des v.a. indépendantes de loi de Bernoulli avec $p_i = \mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0)$ pour tout $1 \leq i \leq n$. Soit μ_n la loi de $S_n = X_1 + \cdots + X_n$ et soit $\nu_n = \mathrm{Poi}(p_1 + \ldots + p_n)$ la loi de Poisson de même moyenne que S_n. Alors on a*

$$d_V(\mu_n, \nu_n) \leq p_1^2 + \cdots p_n^2.$$

Démonstration. On commence par établir par récurrence sur n que si $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ sont des lois de probabilité sur $\mathbb{N}$ alors on a l'inégaltié sous-additive

$$d_V(\alpha_1 * \cdots * \alpha_n, \beta_1 * \cdots * \beta_n) \leq d_V(\alpha_1, \beta_1) + \cdots + d_V(\alpha_n, \beta_n).$$

Ensuite on établit que $d_V(\mathrm{Binom}(1, p), \mathrm{Poi}(p)) \leq p^2$. Rappel : $\mathrm{Poi}(a) * \mathrm{Poi}(b) = \mathrm{Poi}(a + b)$. □

Notons que $p_1^2 + \cdots + p_n^2 \leq (p_1 + \cdots + p_n) \max_{1\leq k\leq n} p_k$.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Annexe A

Compléments

A.1 Lois exponentielles

Les lois exponentielles sont très importantes en modélisation stochastique.

Théorème A.1 (Minimum – Horloges en compétition). *Si $E_1, \ldots, E_n$ sont des v.a.r. indépendantes de loi exponentielle de paramètres $\lambda_1, \ldots, \lambda_n$ alors*

$$M = \min(E_1, \ldots, E_n) \sim \mathrm{Exp}(\lambda_1 + \cdots + \lambda_n).$$

De plus, avec probabilité 1, le minimum M est atteint pour un unique entier aléatoire K indépendant de M et de loi donnée pour tout $1 \leq k \leq n$ par

$$\mathbb{P}(K = k) = \mathbb{P}(M = E_k) = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

Démonstration. On a $\mathbb{P}(M \geq x) = \mathbb{P}(E_1 \geq x) \cdots \mathbb{P}(E_n \geq x) = e^{-(\lambda_1 + \cdots + \lambda_n)x}$ pour tout $x \geq 0$, et cela montre que M suit la loi exponentielle de paramètre $\lambda_1 + \ldots + \lambda_n$. Pour traiter K, nous allons déterminer directement la loi de (M, K), ce qui fournira à nouveau la loi de M. Comme les v.a.r. sont indépendantes et de loi à densité, avec probabilité 1, l'entier aléatoire K est bien défini sur $\{1, \ldots, n\}$. Pour tout $1 \leq k \leq n$ et $t \geq 0$ on a

$$\{M \geq t \text{ et } K = k\} = \{E_k \geq t \text{ et } E_{k'} > E_k \text{ pour tout } k' \neq k\}.$$

Par hypothèse sur les variables aléatoires $E_1, \ldots, E_n$ il vient

$$\mathbb{P}(M \geq t \text{ et } K = k) = \int_t^\infty \lambda_k e^{-s\lambda_k} \prod_{k' \neq k} \mathbb{P}(E_{k'} > s)\, ds = \left(1 - e^{-t(\lambda_1 + \cdots + \lambda_n)}\right) \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

On en déduit que les variables aléatoires M et K sont indépendantes, et de plus, les lois de M et K s'obtiennent en prenant $t = 0$ et en sommant en k respectivement. □

Si $E_1, \ldots, E_n$ sont indépendantes et de loi exponentielle de paramètre λ alors pour tout réel $t \geq 0$ la v.a.r. discrète $\mathbf{1}_{\{E_1 > t\}} + \cdots + \mathbf{1}_{\{E_n > t\}} = \mathrm{Card}\{1 \leq k \leq n : E_k > t\}$ suit la loi binomiale $\mathrm{Binom}(n, e^{-\lambda t})$ car les indicatrices sont indépendantes de loi de Bernoulli.

Théorème A.2 (Comptage). *Soit $(E_n)_{n \geq 1}$ des v.a.r. de loi exponentielle de paramètre λ représentant des durées et pour tout réel $t \geq 0$, la v.a.r. de comptage*

$$N_t = \mathrm{Card}\{n \geq 1 : T_n \leq t\} \quad \text{où} \quad T_n = E_1 + \cdots + E_n.$$

Les propriétés suivantes ont lieu :

Copyrighted material

1. $(T_1, \dots, T_n)$ *a pour densité* $(t_1, \dots, t_n) \mapsto \lambda^n e^{-\lambda t_n} \mathbf{1}_{\{0<t_1<\dots<t_n\}}(t_1, \dots, t_n)$
2. T_n *est la loi Gamma de densité* $t \mapsto \frac{t^{n-1}}{(n-1)!}\lambda^n e^{-\lambda t} \mathbf{1}_{\mathbb{R}_+}(t)$
3. N_t *suit la loi de Poisson* $\mathrm{Poi}(\lambda t)$

On dit que $(N_t)_{t\geq 0}$ est le processus de comptage de la suite $(E_n)_{n\geq 1}$.

Démonstration. La propriété 1 s'obtient par le changement de variable linéaire triangulaire

$$(s_1, s_2, \dots, s_n) \mapsto (s_1, s_1 + s_2, \dots, s_1 + \dots + s_n)$$

à partir de la loi de $(T_1, T_2 - T_1, \dots, T_n - T_{n-1})$ de densité

$$(s_1, \dots, s_n) \mapsto \prod_{i=1}^{n} \lambda e^{-\lambda s_i} \mathbf{1}_{\mathbb{R}_+}(s_i) = \lambda^n e^{-\lambda(s_1+\dots+s_n)} \mathbf{1}_{\mathbb{R}_+^n}(s_1, \dots, s_n).$$

On obtient 2 par récurrence sur n. Pour 3 on écrit $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$, d'où $\mathbb{P}(N_t = n) = \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t)$, et on utilise 2. □

Théorème A.3 (Caractérisation des lois exponentielles par absence de mémoire)**.** *Pour toute v.a.r.* X *sur* $\mathbb{R}_+$ *telle que* $\mathbb{P}(X > 0) > 0$*, les propriétés suivantes sont équivalentes :*

1. $\mathcal{L}(X)$ *est une loi exponentielle ;*
2. $\mathcal{L}(X - t \,|\, X > t) = \mathcal{L}(X)$ *pour tout* $t \geq 0$.

Démonstration. Si $X \sim \mathrm{Exp}(\lambda)$ alors $\mathcal{L}(X - t|X > t) = \mathcal{L}(X)$ car pour tous $s, t \geq 0$,

$$\mathbb{P}(X - t > s|X > t) = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} = e^{-\lambda s}.$$

Réciproquement, si $G(t) = \mathbb{P}(X > t)$, alors la seconde propriété s'écrit : pour tous $s, t \geq 0$

$$G(t)G(s) = G(t + s).$$

Par continuité inférieure, $G(\varepsilon) = \mathbb{P}(X > \varepsilon) \to \mathbb{P}(X > 0) > 0$ lorsque $\varepsilon \to 0^+$. Ainsi, $G(\varepsilon) > 0$ pour $\varepsilon > 0$ assez petit. Si $t > 0$ alors $t \leq n\varepsilon$ avec $n \in \mathbb{N}$ et donc

$$G(t) = \mathbb{P}(X > t) \geq \mathbb{P}(X > n\varepsilon) = G(\varepsilon)^n.$$

Ainsi, $G(t) > 0$ pour tout $t \geq 0$. D'autre part, les solutions non identiquement nulles de l'équation fonctionnelle $G(t+s) = G(t)G(s)$ pour tous $s, t \in \mathbb{R}_+$ sont de la forme $G(t) = G(1)^t$ (considérer les $t \in \mathbb{Q}$ puis utiliser la décroissance de G lorsque $t \in \mathbb{R}_+$). □

Remarque A.4 (Modélisation des durées de vie)**.** *Les lois exponentielles sont utiles pour modéliser les durées dans les processus de renouvellement sans mémoire comme les files d'attente : caisse de supermarché, feu tricolore, etc. En revanche, l'absence de mémoire des lois exponentielles les rend inutilisables pour modéliser des durées de vie : penser à une ampoule électrique. On modélise une durée de vie par une variable aléatoire réelle positive* X*. On peut penser à la durée de vie d'un organisme vivant, d'une entreprise, d'une situation donnée comme le chômage, etc. Supposons que* X *admet une densité* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *telle que* $f(t) > 0$ *pour tout* $t > 0$*. La loi de* X *est caractérisée par la donnée de la densité* f *ou encore par la donnée de la fonction de répartition* F *définie par*

$$F(t) = \mathbb{P}(X \leq t) = \int_0^t f(s)\, ds$$

pour tout $t \geq 0$*. La* fonction de survie *est donnée par* $S(t) = 1 - F(t) = \mathbb{P}(X > t)$ *pour tout* $t \geq 0$*. Elle représente la probabilité de mourir après l'instant* t*, i.e. de survivre à la période* $[0, t]$*. Il est clair que la fonction de survie caractérise la loi de* X*, car il en va de même de la fonction de répartition* F*. La fonction de survie vérifie :*

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

- S prend ses valeurs dans l'intervalle $[0, 1]$
- S est continue
- S est décroissante
- $S(0) = 1$ et $\lim_{t\to\infty} S(t) = 0$

La fonction de hasard h est définie pour tout $t \geq 0$ par

$$h(t) = \frac{f(t)}{S(t)} = \frac{F'(t)}{1 - F(t)} = -\frac{S'(t)}{S(t)}.$$

Le terme hasard est un anglicisme qui signifie ici danger ou risque. La fonction de hasard est positive. Elle s'interprète comme un taux de sortie instantané : pour tout $t \geq 0$,

$$h(t) = \frac{d}{ds}\bigg|_{s=0} \mathbb{P}(t < X < t + s \,|\, X > t).$$

Pour le voir, on écrit

$$\begin{aligned}
\frac{F(t+s)}{S(t)} &= \frac{\mathbb{P}(X < t + s)}{\mathbb{P}(X > t)} \\
&= \frac{\mathbb{P}(X < t) + \mathbb{P}(t < X < t + s)}{\mathbb{P}(X > t)} \\
&= 1 + \frac{\mathbb{P}(t < X < t + s)}{\mathbb{P}(X > t)} \\
&= 1 + \mathbb{P}(t < X < t + s \,|\, X > t).
\end{aligned}$$

La fonction de hasard caractérise la loi. En effet, pour tout $t \geq 0$, on a

$$S(t) = \exp\left(-\int_0^t h(u)\, du\right).$$

Pour tout $t_0 \geq 0$, la fonction de survie conditionnelle est définie pour tout $t > 0$ par

$$S(t \,|\, t_0) = \mathbb{P}(X > t + t_0 \,|\, X > t_0).$$

Les fonctions de survie conditionnelles caractérisent la loi car pour tout $t_0, t > 0$,

$$S(t \,|\, t_0) = \frac{S(t + t_0)}{S(t_0)} = \exp\left(-\int_{t_0}^{t+t_0} h(s)\, ds\right).$$

On cherche souvent à modéliser les durées de vies via leur fonction de hasard h plutôt que via la densité f. Voici deux exemples concrets de fonctions de hasard :

- Durée de vie humaine : la fonction h part d'une valeur positive et décroît violemment (ce pic initial correspond à la mortalité infantile), puis a un plat muni de deux bosses vers 18–22 ans (accidents deux roues) et 40 ans (accidents cardiaques), puis remonte de manière convexe (vieillesse). Les pics de la fonctions de hasard correspondent à une diminution de la durée de vie.
- Durée du chômage : la fonction h a l'allure de la fonction $x \mapsto (1 + x)exp(-x)$, le pic mou correspondant au chômage de longue durée.

Supposons X intégrable. La durée de vie moyenne restante est définie pour $t \geq 0$ par

$$r(t) = \mathbb{E}(X - t \,|\, X > t) = \mathbb{E}(X - t \,|\, X > t) - t.$$

Cette fonction caractérise la loi de X. En effet, il est clair que r est entièrement déterminée par la loi de X. Réciproquement, connaissant la fonction r, on a pour tout $t \geq 0$,

$$r(t) = \frac{1}{S(t)} \int_t^\infty (s - t)\, f(s)\, ds.$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Or par intégration par parties, on a

$$\int_t^\infty (s-t)\, f(s)\, ds = -\int_t^\infty (s-t)\, S'(s)\, ds = -[(s-t)S(s)]_t^\infty + \int_t^\infty S(s)\, ds.$$

Le terme de bord est nul car X est intégrable et on obtient

$$\frac{1}{r(t)} = \frac{S(t)}{\displaystyle\int_t^\infty S(s)\, ds} = -\frac{d}{dt}\ln\left(\int_t^\infty S(s)\, ds\right).$$

Cela donne

$$\ln\int_t^\infty S(s)\, ds - \ln\int_0^\infty S(s)\, ds = -\int_0^t \frac{1}{r(s)}\, ds$$

d'où, en remarquant que $-\displaystyle\int_0^\infty S(s)\, ds = r(0)$ ($= \mathbb{E}(X)$),

$$\int_t^\infty S(s)\, ds = r(0)\exp\left(-\int_0^t \frac{1}{r(s)}\, ds\right).$$

Par conséquent, en dérivant en t, on obtient,

$$S(t) = \frac{r(0)}{r(t)}\exp\left(-\int_0^t \frac{1}{r(s)}\, ds\right).$$

Ainsi, la durée de vie moyenne restante caractérise bien la loi. On dispose d'une caractérisation de l'indépendance temporelle qui découle de l'absence de mémoire des lois exponentielles : pour tout $\lambda > 0$, il y a équivalence entre :

- *X suit la loi exponentielle de moyenne $1/\lambda$*
- *$f(t) = \lambda e^{-\lambda t}$ pour tout $t \geq 0$*
- *$F(t) = 1 - e^{-\lambda t}$ pour tout $t \geq 0$*
- *$S(t) = e^{-\lambda t}$ pour tout $t \geq 0$*
- *$h(t) = \lambda$ pour tout $t \geq 0$ (constance de la fonction de hasard)*
- *$S(t \mid t_0) = e^{-\lambda t}$ pour tous $t_0, t \geq 0$ (absence de mémoire)*
- *$S(t + t_0) = S(t)S(t_0)$ pour tous $t_0, t \geq 0$*
- *$r(t) = 1/\lambda$ pour tout $t \geq 0$*

Notons que la durée de vie d'un montage de composants en série correspond à un calcul de minimum tandis que la durée de vie d'un montage en parallèle correspond à un calcul de maximum. On a coutume de dire que la fonction de hasard typique d'un composant a une forme de baignoire : rodage (décroissance), exploitation (long plateau), usure (croissance).

Exercice A.5 (Fiabilité des systèmes)**.** *L'étude quantitative de la fiabilité des systèmes fait appel aux concepts de fonction de survie et de fonction de hasard. À ce sujet, il est instructif d'effectuer une recherche sur Internet sur les termes suivants : New Better Than Used, Mean Time To Failure, fonction de structure, coupes minimales, arbre de défaillance.*

Théorème A.6 (Lois exponentielles et géométriques)**.** *La contraction des géométriques donne des exponentielles et la discrétisation des exponentielles donne des géométriques :*

1. *Contraction de la loi géométrique. Pour tout $n \geq 1$ soit $X_n \sim \mathrm{Geom}_{\mathbb{N}^*}(p_n)$ avec $0 < p_n < 1$. Si $\lim_{n\to\infty} np_n = \lambda > 0$ alors pour tout $x \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_n}{n} \geq x\right) = e^{-\lambda x}.$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

2. *Discrétisation de la loi exponentielle. Si $Y \sim \mathrm{Exp}(\lambda)$ et si $\lfloor Y \rfloor$ désigne la partie entière de Y, alors $\lfloor Y \rfloor$ et $Y - \lfloor Y \rfloor$ sont indépendantes et $1 + \lfloor Y \rfloor \sim \mathrm{Geom}_{\mathbb{N}^*}(e^{-\lambda})$.*

Démonstration. Pour la première partie on écrit, pour tout $n \geq 1$ et tout $x \in \mathbb{R}$,

$$\mathbb{P}(X_n \geq nx) = \mathbb{P}(X_n \geq \lfloor nx \rfloor) = (1 - p_n)^{\lfloor nx \rfloor} \to e^{-\lambda x}.$$

La seconde partie découle de la formule $\mathbb{P}(Y > t + s \,|\, Y > s) = \mathbb{P}(Y > t) = e^{-\lambda t}$ pour tous $s, t \geq 0$, qui correspond à l'absence de mémoire des lois exponentielles. □

Théorème A.7 (Maximum). *Si $E_1, \ldots, E_n$ sont indépendantes de loi $\mathrm{Exp}(\lambda)$ alors*

$$\mathcal{L}(\max(E_1, \ldots, E_n)) = \mathcal{L}(F_1 + \cdots + F_n)$$

où $F_1, \ldots, F_n$ sont indépendantes avec $F_k \sim \mathrm{Exp}(k\lambda)$ pour tout $1 \leq k \leq n$.

Démonstration. Posons $S_n = F_1 + \cdots + F_n$. La densité de $M_n = \max(E_1, \ldots, E_n)$ est

$$f_n(x) = (\mathbb{P}(M_n \leq x))' = ((1 - e^{-\lambda x})^n \mathbf{1}_{\mathbb{R}_+}(x))' = n\lambda(1 - e^{-\lambda x})^{n-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Montrons par récurrence sur n que S_n a pour densité f_n. C'est vrai pour $n = 1$. Si cela est vrai pour n, alors la densité de S_{n+1} est, en notant g_λ la densité de $\mathrm{Exp}(\lambda)$,

$$\begin{aligned} f_n * g_{(n+1)\lambda}(y) &= (n+1)\lambda \int_{-\infty}^{y} f_n(x) e^{-\lambda(n+1)(y-x)}\, dx \\ &= \lambda(n+1) n\lambda e^{-\lambda(n+1)y} \int_0^y (e^{\lambda x} - 1)^{n-1} e^{\lambda x}\, dx \\ &= \lambda(n+1) e^{-\lambda(n+1)y} (e^{\lambda y} - 1)^n \\ &= \lambda(n+1) e^{-\lambda y} (1 - e^{-n\lambda y})^n = f_{n+1}(y). \end{aligned}$$

□

Considérons un arbre binaire infini représentant la descendance d'une cellule mère. Supposons que chaque bout de branche a une longueur aléatoire qui représente la durée de vie avant division. On suppose que toutes ces longueurs sont des v.a.r. indépendantes de loi exponentielle de paramètre λ. À l'instant $t \geq 0$, l'arbre possède N_t branches, et $N_0 = 1$. Le processus $(N_t)_{t \geq 0}$ est appelé processus de Yule de paramètre λ.

Corollaire A.8 (Taille de la population du processus de Yule). *Pour tout $t \geq 0$ la taille de la population N_t suit la loi géométrique sur $\mathbb{N}^*$ de paramètre $e^{-\lambda t}$. En particulier,*

$$\mathbb{E}(N_t) = e^{\lambda t} \quad \textit{et} \quad \mathrm{Var}(N_t) = (1 - e^{-\lambda t}) e^{2\lambda t}.$$

Démonstration. La propriété d'absence de mémoire des lois exponentielles entraîne que pour tout $t \geq 0$, on a $N_t = \mathrm{Card}\{n \geq 1 : S_n \leq t\}$ où $S_n = F_1 + \cdots + F_n$ où les $(F_n)_{n \geq 1}$ sont des v.a.r. indépendantes avec $F_n \sim \mathrm{Exp}(n\lambda)$ pour tout $n \geq 1$. Or d'après le théorème A.7, la v.a.r. S_n a la loi de $\max(E_1, \ldots, E_n)$ où $E_1, \ldots, E_n$ sont des v.a.r. indépendantes de loi exponentielle de paramètre λ. Donc, pour tout $n \geq 0$,

$$\mathbb{P}(N_t - 1 \geq n) = \mathbb{P}(S_n \leq t) = \mathbb{P}(E_1 \leq t) \cdots \mathbb{P}(E_n \leq t) = (1 - e^{-\lambda t})^n.$$

□

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

A.2 Jeu de pile ou face

On modélise une infinité de lancers au jeu de pile ou face avec une pièce équilibrée par l'espace probabilisé produit $(\{0,1\}^{\mathbb{N}^*}, \mathcal{F}, \mathbb{P})$ où $\mathcal{F}$ est la tribu engendrée par les cylindres et $\mathbb{P}$ la mesure de probabilité produit associée à la mesure de probabilité uniforme sur $\{0,1\}$. Ici pile est codé 1 et face est codé 0. Lorsque la pièce n'est pas équilibrée et donne face avec probabilité $p \in [0,1]$, on équipe $\{0,1\}$ de la loi qui affecte la probabilité p à 1 et $1-p$ à 0. La suite des coordonnées dans cet espace produit constitue une suite de v.a.r. $(X_n)_{n\geq 1}$ indépendantes et de même loi de Bernoulli :

$$\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = p \in [0,1].$$

Nombre de succès en n lancers

Le nombre de succès dans les n premiers lancers $S_n = X_1 + \cdots + X_n$ suit la loi binomiale $\mathrm{Binom}(n,p)$ de taille n et de paramètre p, donnée pour tout $k = 0, 1, \ldots, n$ par

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

On a par linéarité de l'espérance et indépendance des $X_1, \ldots, X_n$

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np \quad \text{et} \quad \sigma^2(S_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n) = np(1-p).$$

Temps d'attente du premier succès

Si $p > 0$, alors le nombre de lancers pour obtenir un succès $T = \inf\{n \geq 1 : X_n = 1\}$ suit la loi géométrique $\mathrm{Geo}_{\mathbb{N}^*}(p)$ sur $\mathbb{N}^*$ de paramètre p donnée pour tout $k \in \mathbb{N}^*$

$$\mathbb{P}(T = k) = (1-p)^{k-1} p.$$

On a $T \equiv \infty$ si $p = 0$ et $\mathbb{P}(T < \infty) = 1$ sinon. On a

$$\mathbb{E}(T) = \frac{1}{p} \quad \text{et} \quad \sigma^2(T) = \frac{1-p}{p^2}.$$

Le nombre d'échecs avant le premier succès $T' = \inf\{n \geq 0 : X_{n+1} = 1\} = T - 1$ suit la loi géométrique $\mathrm{Geo}_{\mathbb{N}}(p)$ sur $\mathbb{N}$ et de paramètre p donnée pour tout $k \in \mathbb{N}$ par

$$\mathbb{P}(T' = k) = \mathbb{P}(T - 1 = k) = (1-p)^k p$$

et on a

$$\mathbb{E}(T') = \mathbb{E}(T) - 1 = \frac{1-p}{p} \quad \text{et} \quad \sigma^2(T') = \sigma^2(T) = \frac{1-p}{p^2}.$$

Temps d'attente des succès successifs

Pour tout $r \in \mathbb{N}^*$, le nombre de lancers T_r nécessaires pour obtenir r succès est défini par récurrence par $T_1 = T$ et $T_{r+1} = \inf\{n > T_r : X_n = 1\}$. Les v.a.r. $T_1, T_2 - T_1, T_3 - T_2, \ldots$ sont indépendantes et de même loi géométrique $\mathrm{Geo}_{\mathbb{N}^*}(p)$. La variable aléatoire T_r suit la loi de Pascal ou loi binomiale-négative $\mathrm{Geo}_{\mathbb{N}^*}(p)^{*r}$. On a pour tout $k \geq r$,

$$\mathbb{P}(T_r = k) = \sum_{\substack{k_1 \geq 1, \ldots, k_r \geq 1 \\ k_1 + \cdots + k_r = k}} (1-p)^{k_1 - 1} p \cdots (1-p)^{k_r - 1} p = (1-p)^{k-r} p^r \binom{k-1}{r-1}$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

et

$$\mathbb{E}(T_r) = r\mathbb{E}(T) = \frac{r}{p} \quad \text{et} \quad \sigma^2(T_r) = r\sigma^2(T) = r\frac{1-p}{p^2}.$$

Le processus de Bernoulli $(B_n)_{n\geq 0}$ est donné par $B_n = B_0 + S_n$ où B_0 est une variable aléatoire quelconque. Ses trajectoires sont constantes par morceaux, avec des sauts d'amplitude $+1$, et les temps de saut sont donnés par $(T_r)_{r\geq 1}$ (temps inter-sauts indépendantes et de même loi géométrique). Le processus de Bernoulli est le processus de comptage de tops espacés par des durées indépendantes de même loi géométriques. De ce point de vue, il constitue un analogue à temps discret du processus de Poisson simple.

Théorèmes limites

La loi forte des grands nombres et le théorème de la limite centrale s'écrivent ici

$$\frac{S_n}{n} \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} = p \quad \text{et} \quad \frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\frac{S_n}{n} - p\right) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1).$$

Cela donne un intervalle de confiance asymptotique pour p appelé intervalle de Wald (assez mauvais en pratique). Il est également possible de confectionner des intervalles de confiance pour p non asymptotiques, comme celui de Clopper-Pearson par exemple, basé sur la correspondance beta-binomiale : si $U_1, \ldots, U_n$ sont des v.a. indépendantes et de loi uniforme sur $[0,1]$ et si $U_{(1)} \leq \cdots \leq U_{(n)}$ est leur réordonnement alors $U_{(k)} \sim \mathrm{Beta}(k, n-k+1)$ a pour densité $t \in [0,1] \mapsto (\int_0^1 s^{k-1}(1-s)^{n-k}\,ds)^{-1} t^{k-1}(1-t)^{n-k}$, et

$$\mathbb{P}(S_n \geq k) = \mathbb{P}(\mathbf{1}_{\{U_1\leq p\}} + \cdots + \mathbf{1}_{\{U_n\leq p\}} \geq k) = \mathbb{P}(U_{(k)} \leq p).$$

Motifs répétés

Si $A_n =$ «le n^e lancer est pile» $= \{X_n = 1\}$ alors $\mathbb{P}(A_n) = p$ et donc $\sum_n \mathbb{P}(A_n) = \infty$ lorsque $p > 0$. Comme les (A_n) sont indépendants, la loi du zéro-un de Borel (seconde partie du Borel-Cantelli) implique alors que $\mathbb{P}(\overline{\lim}_n A_n) = 1$, ce qui signifie qu'avec probabilité 1, on obtient «pile» une infinité de fois. Idem pour «face» lorsque $p < 1$. Plus généralement, si $\varepsilon_1 \ldots \varepsilon_r$ est un mot binaire fixé de longueur r construit avec $\varepsilon_1, \ldots, \varepsilon_r \in \{0,1\}$, et si $0 < p < 1$, alors avec probabilité 1, ce mot binaire apparaît une infinité de fois dans le jeu de pile ou face (considérer le v.a.r. indépendantes et de même loi $Y_1 = (X_1, \ldots, X_r), Y_2 = (X_{r+1}, \ldots, X_{2r}), \ldots$. Ainsi, si T est la traduction en une suite finie de 0 et de 1 de la Bible ou de Hamlet, alors avec probabilité 1, la suite T apparaît une infinité de fois à pile ou face (le fameux singe et sa machine à écrire...).

Lien avec la loi uniforme sur $[0,1]$

Chaque réalisation de la suite (X_n) permet de construire un nombre réel dans l'intervalle $[0,1]$ via son écriture en base 2. Cela correspond à la surjection $\varsigma : \{0,\ldots,1\}^{\mathbb{N}^*} \to [0,1]$ définie pour tout $x \in \{0,\ldots,2\}^{\mathbb{N}^*}$ par

$$\varsigma(x) = \sum_{n=1}^{\infty} 2^{-n} x_n = \underbrace{0, x_1 \cdots x_n \cdots}_{\text{en base } 2}.$$

Lorsque $p = 1/2$ alors la variable aléatoire U donnée par

$$U = \varsigma(X) = \sum_{n=1}^{\infty} 2^{-n} X_n$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

suit la loi uniforme sur $[0,1]$. Si $a = a_1 r^{-1} + \cdots + a_n 2^{-n}$ est un nombre diadique, on a

$$\begin{aligned}\mathbb{P}(a < U < a + 2^{-n}) &= \mathbb{P}(X_1 = a_1, \ldots, X_n = a_n) \\ &= \mathbb{P}(X_1 = a_1) \cdots \mathbb{P}(X_n = a_n) = 2^{-n}.\end{aligned}$$

Réciproquement, ce calcul montre que les coefficients (X_n) de l'écriture en base 2 d'une variable aléatoire uniforme sur $[0,1]$ sont indépendants de loi de Bernoulli symétrique sur $\{0,1\}$. Les nombres de $[0,1]$ dont l'écriture en base 2 est constante à partir d'un certain rang « ne comptent pas » en quelque sorte, et ce est presque sûrement une injection et donc presque sûrement une bijection. On peut donc en déduire une méthode pour générer d'un seul coup n réalisations indépendantes de loi de Bernoulli symétrique sur $\{0,1\}$ à partir d'une réalisation de précision n en base 2 d'une loi uniforme sur $[0,1]$.

A.3 Collectionneur de coupons

Le collectionneur de coupons constitue un modèle stochastique fondamental important, à ranger dans la même boite à outils que le jeu de pile ou face, auquel il est intimement relié. Un grand nombre de situations concrètes sont modélisables par le collectionneur de coupons ou une de ses variantes. Nous nous limitons ici à la variante la plus simple.

Il faut jouer un nombre de fois (aléatoire) géométrique à pile ou face pour voir apparaître les deux côtés de la pièce. Si on remplace la pièce de monnaie par un dé à $r \geq 2$ faces, combien de fois faut-il lancer le dé pour voir apparaître les r faces différentes ? On modélise cela, pour un entier fixé $r \geq 2$, en considérant la variable aléatoire

$$T = \min\{n \geq 1 : \{X_1, \ldots, X_n\} = \{1, \ldots, r\}\} = \min\{n \geq 1 : \mathrm{Card}\{X_1, \ldots, X_n\} = r\}$$

où $(X_n)_{n\geq 1}$ est une suite de v.a.r. indépendantes et de même loi uniforme sur $\{1, \ldots, r\}$. La variable aléatoire T est le temps de complétion de la collection. Le nom *collectionneur de coupons* provient des coupons à collectionner présents dans certains paquets de céréales.

Théorème A.9 (Expression combinatoire de la loi). *On a $T \geq r$ et pour tout $n \geq r$,*

$$\mathbb{P}(T = n) = \frac{r!}{r^n} \left\{ \begin{matrix} n-1 \\ r-1 \end{matrix} \right\}$$

où la notation entre accolades est le nombre de Stirling de seconde espèce $(n-1, r-1)$, c'est-à-dire le nombre de partitions en $r-1$ blocs d'un ensemble de $n-1$ éléments.

Démonstration. On a $X_T \notin \{X_1, \ldots, X_{T-1}\}$ car le coupon qui termine la collection n'a forcément jamais été vu auparavant. Si on fixe $n \geq r$, l'évènement $\{T = n\}$ correspond à choisir le type du dernier coupon puis à répartir les $n-1$ coupons restants sur les $r-1$ types restants. Le résultat désiré en découle car la loi des type est uniforme. □

Le superbe théorème A.9 n'est malgré tout pas très parlant. Le résultat intuitif suivant va beaucoup nous aider à étudier T, et montre en particulier que $\mathbb{P}(T < \infty) = 1$.

Lemme A.10 (Décomposition). *On a $T = G_1 + \cdots + G_r$ où $G_1, \ldots, G_r$ sont des v.a. indépendantes, géométriques sur $\{1, 2, \ldots\}$, avec, pour tout $1 \leq i \leq r$, G_i de paramètre*

$$\pi_i = \frac{r-i+1}{r}.$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Démonstration. On pose $G_1 \equiv 1$ et pour tout $1 < i \leq r$,

$$G_i = \min\{n \geq 1 : X_{G_{i-1}+n} \notin \{X_1, \ldots, X_{G_{i-1}}\}\}.$$

On a $\mathrm{Card}(\{X_1, \ldots, X_{G_i}\}) = i$ pour tout $1 \leq i \leq n$. Les variables aléatoires $G_1, G_1 + G_2, \ldots, G_1 + \cdots + G_r$ sont les temps d'apparition des r premiers gains dans un jeu de pile ou face spécial dans lequel la probabilité de gagner change après chaque gain : cette probabilité vaut successivement $1, (r-1)/r, (r-2)/r, \ldots, 1/r$. Cela témoigne du fait qu'il est de plus en plus difficile d'obtenir un coupon d'un nouveau type au fil de la collection. □

Théorème A.11 (Queue de distribution). *Pour tout $n \geq 1$,*

$$\mathbb{P}(T > n) = \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^n.$$

Démonstration. On a

$$\mathbb{P}(T > n) = \mathbb{P}(E_{n,1} \cup \cdots \cup E_{n,r}) \quad \text{où} \quad \mathbb{P}(E_{n,i}) = \{X_1 \neq i, \ldots, X_n \neq i\}.$$

Si $i_1, \ldots, i_k \in \{1, \ldots, r\}$ sont distincts alors, avec $R = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$,

$$\mathbb{P}(E_{n,i_1} \cap \cdots \cap E_{n,i_k}) = \mathbb{P}(X_1 \in R) \cdots \mathbb{P}(X_n \in R) = \left(\frac{r-k}{r}\right)^n = \left(1 - \frac{k}{r}\right)^n.$$

Le résultat désiré découle alors du principe d'inclusion-exclusion (théorème 2.17). □

Dans la formule du théorème A.11, les signes sont alternés, ce qui rend délicat d'en déduire le comportement de la queue de T en fonction de n, r.

Théorème A.12 (Déviation). *Pour tout réel $t > 0$,*

$$\mathbb{P}(T > 1 + \lfloor tr \ln(r) \rfloor) \leq r^{-t+1}.$$

Démonstration. Pour tout entier $n \geq 1$, on peut écrire

$$\mathbb{P}(T > n) = \mathbb{P}(\cup_{i=1}^{r} E_{n,i}) \leq \sum_{i=1}^{r} \mathbb{P}(E_{n,i}) \quad \text{où} \quad E_{n,i} = \{X_1 \neq i, \ldots, X_n \neq i\}.$$

Comme $\mathbb{P}(E_{n,i}) = (1 - 1/r)^n \leq e^{-n/r}$, le choix $n = 1 + \lfloor tr \ln(r) \rfloor$ mène au résultat. □

À présent, pour $\alpha = 0.05$ et r fixé, on peut choisir t assez grand pour que $r^{-t+1} \leq \alpha$, ce qui fournit l'intervalle de prédiction $[r, \lfloor tr \ln(r) \rfloor + 1]$ de niveau α.

Par le lemme A.10 et la linéarité de l'espérance, on a

$$\mathbb{E}(T) = \sum_{i=1}^{r} \mathbb{E}(G_i) = \sum_{i=1}^{r} \frac{1}{\pi_i} = \sum_{i=1}^{r} \frac{r}{r-i+1} = r \sum_{i=1}^{r} \frac{1}{i} = r(\ln(r) + \gamma + o_{r\to\infty}(1))$$

où $\gamma = \lim_{n\to\infty}(\sum_{i=1}^{n} 1/i - \ln(n)) \approx 0.577$ est la constante d'Euler. Comme les $G_1, \ldots, G_r$ sont indépendantes avec $\sigma^2(G_i) = (1 - \pi_i)/\pi_i^2 = r(i-1)/(r-i+1)^2$, on a également

$$\sigma^2(T) = \sum_{i=1}^{r} \sigma^2(G_i) = r \sum_{i=1}^{r-1} \frac{r-i}{i^2} = \frac{\pi^2}{6} r^2 - r \ln(r) - r + o_{r\to\infty} r^2.$$

Théorème A.13 (Convergence).

$$\frac{T}{r \ln(r)} \xrightarrow[r\to\infty]{\mathbb{P}} 1.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. Par l'inégalité de Markov et les formules pour les moments de T, il vient

$$\begin{aligned}\mathbb{P}\left(\left|\frac{T}{r\ln(r)}-1\right|>t\right) &\leq \frac{\mathbb{E}(|T-r\ln(r)|^2)}{r^2\ln(r)^2t^2}\\ &= \frac{\sigma^2(T)+(\mathbb{E}(T)-r\ln(r))^2}{r^2\ln(r)^2t^2}\\ &= \mathcal{O}_{r\to\infty}\left(\frac{1}{\ln(r)^2}\right).\end{aligned}$$

□

La borne logarithmique sur la vitesse de convergence en probabilité est trop faible pour en déduire une convergence presque sûre au moyen du lemme de Borel-Cantelli. En revanche, la majoration obtenue par l'inégalité de Markov fournit un intervalle de prédiction pour T : pour tout $t>0$,

$$\mathbb{P}(T\in[r\ln(r)-rt, r\ln(r)+rt])\geq 1-\frac{\mathcal{O}_{r\to\infty}(1)}{t^2}.$$

À présent, pour $\alpha=0.05$ et r fixé, il faut choisir t assez grand pour que le second membre soit égal à $1-\alpha$. L'intervalle de prédiction, de largeur $2rt$, se dégrade quand t augmente.

On rappelle que la loi de Gumbel a pour fonction de répartition $t\in\mathbb{R}\mapsto e^{-e^{-t}}$.

Théorème A.14 (Fluctuations asymptotiques). *On a*

$$\frac{T-r\ln(r)}{r}\xrightarrow[r\to\infty]{\text{loi}}\text{Gumbel}.$$

Démonstration. Il suffit d'établir que pour tout $t\in\mathbb{R}$ on a

$$\lim_{r\to\infty}\mathbb{P}(T>r\ln(r)+tr)=S(t)=1-e^{-e^{-t}}.$$

Fixons donc $t\in\mathbb{R}$ et supposons que r est assez grand pour que $r\ln(r)+tr>r$. Soit $n_{t,r}$ l'entier défini par $n_{t,r}=r\ln(r)+tr$ si $r\ln(r)+tr\in\mathbb{N}$ et $n_{t,r}=\lfloor r\ln(r)+tr\rfloor+1$ sinon. Le théorème A.11 donne

$$\mathbb{P}(T>r\ln(r)+tr)=\sum_{k=1}^{r}(-1)^{k-1}\binom{r}{k}\left(1-\frac{k}{r}\right)^{n_{t,r}}.$$

Comme $\binom{r}{k}\leq r^k/k!$ et $1-u\leq e^{-u}$ pour tout $u\geq 0$, on a

$$\binom{r}{k}\left(1-\frac{k}{r}\right)^{n_{t,r}}\underset{r\to\infty}{\leq}\frac{e^{-tk}}{k!}.$$

Enfin, par convergence dominée, on obtient

$$\lim_{r\to\infty}\sum_{k=1}^{r}(-1)^{k-1}\binom{r}{k}\left(1-\frac{k}{r}\right)^{n_{t,r}}=\sum_{k=1}^{\infty}(-1)^{k-1}\frac{e^{-tk}}{k!}=S(t).$$

□

Le théorème A.14 fournit un intervalle de prédiction pour T : pour tout réel $t\geq 0$,

$$\lim_{r\to\infty}\mathbb{P}(T\in[r\ln(r)-rt, r\ln(r)+rt])=e^{-e^{-t}}-e^{-e^{t}}.$$

Lorsque r est inconnu, un estimateur naturel est $\widehat{r}_n=\max\{X_1,\ldots,X_n\}=\text{Card}\{X_1,\ldots,X_n\}$. Il constitue un analogue discret de l'estimateur $\widehat{\theta}_n=\max\{U_1,\ldots,U_n\}$ de θ où $U_1,\ldots,U_n$ sont i.i.d. de loi uniforme sur $[0,\theta]$.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

A.4 Marche aléatoire simple et ruine du joueur

Soit $0 < p < 1$ un réel fixé et soit $(\varepsilon_n)_{n\geq 1}$ une suite de v.a. indépendantes et de même loi telles que $\mathbb{P}(\varepsilon_n = 1) = 1 - \mathbb{P}(\varepsilon_n = -1) = p$ pour tout $n \geq 1$. Soit X_0 une v.a. sur $\mathbb{Z}$ indépendante de la suite $(\varepsilon_n)_{n\geq 1}$. La *marche aléatoire simple* sur $\mathbb{Z}$, de paramètre p, est la suite récurrence aléatoire $(X_n)_{n\geq 0}$ sur $\mathbb{Z}$ définie par relation récursive suivante :

$$X_{n+1} = X_n + \varepsilon_{n+1} = X_0 + \varepsilon_1 + \cdots + \varepsilon_{n+1}$$

pour tout $n \geq 0$. Pour tout $n \geq 1$, la v.a. $\beta_n = (\varepsilon_n + 1)/2$ suit la loi de Bernoulli de paramètre p car on a $\mathbb{P}(\beta_n = 1) = 1 - \mathbb{P}(\beta_n = 0) = p$. Ainsi, pour tout $n \geq 0$,

$$\frac{X_n - X_0 + n}{2} = \beta_1 + \cdots + \beta_n \sim \mathrm{Binom}(n, p).$$

Le théorème suivant permet d'étudier le problème de la ruine d'un joueur qui gagne 1 Euro avec probabilité p et perd 1 Euro avec probabilité $1 - p$. La fortune initiale est x et le joueur quitte le jeu lorsqu'il possède $a < x$ Euros (ruine) ou $b > x$ Euros (gain).

Théorème A.15 (Sortie ou ruine). *Soient $a, b \in \mathbb{Z}$ avec $a < b$. Soit τ_a, τ_b et τ les v.a. à valeurs dans $\mathbb{N} \cup \{\infty\}$ définies par*

$$\tau_a = \inf\{n \geq 0 : X_n = a\}, \quad \tau_b = \inf\{n \geq 0 : X_n = b\}, \quad \textit{et} \quad \tau = \min(\tau_a, \tau_b).$$

Alors pour tout $a \leq x \leq b$ on a $\mathbb{E}_x(\tau) < \infty$ et en posant $\rho = \dfrac{1-p}{p}$ on a

$$\mathbb{P}_x(X_\tau = a) = \begin{cases} \dfrac{\rho^b - \rho^x}{\rho^b - \rho^a} & \textit{si } p \neq \frac{1}{2}, \\ \dfrac{b-x}{b-a} & \textit{si } p = \frac{1}{2}. \end{cases} \quad \textit{et} \quad \mathbb{E}_x(\tau) = \begin{cases} \dfrac{x-a}{1-2p} - \dfrac{(b-a)}{1-2p}\dfrac{\rho^x - \rho^a}{\rho^b - \rho^a} & \textit{si } p \neq \frac{1}{2}, \\ (b-x)(x-a) & \textit{si } p = \frac{1}{2}. \end{cases}$$

Démonstration. Montrons que $\mathbb{E}_x(\tau) < \infty$ pour tout $a \leq x \leq b$. Pour tout $a \leq x \leq b$, il existe un chemin ℓ_x de longueur $|\ell_x| \leq (b-a)$ qui mène de x à a ou b. On a

$$\mathbb{P}_x(\tau > (b-a)) \leq \mathbb{P}(X_{1:|\ell_x|} \neq \ell_x) = 1 - \mathbb{P}(X_{1:|\ell_x|} = \ell_x) \leq 1 - \min(p, 1-p)^{|\ell_x|}.$$

Si $\eta = \max_{a<x<b}(1 - \min(p, 1-p))^{|\ell_x|}$ alors on obtient pour tout $k \geq 1$,

$$\mathbb{P}_x(\tau > k(b-a)) \leq \eta^k.$$

Comme $\eta < 1$ on obtient que $\mathbb{E}_x(\tau) < \infty$ (tous les moments sont finis en fait) et en particulier $\mathbb{P}_x(\tau < \infty) = 1$. Calculons $r(x) = \mathbb{P}_x(X_\tau = a)$. On a pour tout $a < x < b$

$$\begin{aligned} r(x) &= \mathbb{P}_x(X_\tau = a \mid X_1 = x+1)p + \mathbb{P}_x(X_\tau = a \mid X_1 = x-1)(1-p) \\ &= pr(x+1) + (1-p)r(x-1). \end{aligned}$$

L'ensemble des solutions de cette récurrence linéaire d'ordre deux est un espace vectoriel qui contient la solution constante 1. Si $p \neq 1/2$ alors ρ^x est aussi solution, linéairement indépendante de 1, et donc les solutions sont de la forme $A + B\rho^x$ avec A et B constantes. Les conditions aux bords $r(a) = 1$, $r(b) = 0$ fixent A et B, ce qui donne l'unique solution

$$r(x) = \frac{\rho^b - \rho^x}{\rho^b - \rho^a}.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Si $p = 1/2$ alors $\rho = 1$ et les deux solutions fondamentales précédentes sont confondues. Cependant, on observe que dans ce cas, x est également solution, linéairement indépendante de 1, et donc les solutions sont de la forme $A + Bx$ où A et B sont des constantes. Les conditions aux bords $r(a) = 1$ et $r(b) = 0$ fixent A et B, ce qui donne l'unique solution

$$r(x) = \frac{b-x}{b-a}.$$

Calculons $R(x) = \mathbb{E}_x(\tau)$. En conditionnant selon X_1 on obtient pour tout $a < x < b$ la récurrence linéaire (la méthode est valable pour toute chaîne de Markov, idem pour $r(x)$)

$$R(x) = pR(x+1) + (1-p)R(x-1) + 1.$$

La présence du second membre 1 fait rechercher des solutions particulières. Si $p \neq 1/2$ alors $x/(1-2p)$ est solution particulière, et les solutions de l'équation sont de la forme $R(x) = x/(1-2p) + A + B\rho^x$. Les conditions aux bords $R(a) = 0$ et $R(b) = 0$ donnent enfin

$$R(x) = \frac{x-a}{1-2p} - \frac{(b-a)}{1-2p}\frac{\rho^b - \rho^x}{\rho^b - \rho^a}.$$

Si $p = 1/2$ alors $-x^2$ est solution particulière, et les solutions sont de la forme $-x^2 + A + Bx$. Les conditions aux bords $R(a) = R(b) = 0$ donnent enfin

$$R(x) = (b-x)(x-a).$$

La même approche permet de calculer la fonction génératrice $\mathbb{E}_x(s^\tau \mid X_\tau = a)$. □

Remarque A.16. *Si $p = 1/2$, on a $\mathbb{P}_x(\tau_a < \infty) = 1$ et $\mathbb{P}_x(\tau_b < \infty) = 1$ pour tout $a \leq x \leq b$. Il est possible d'établir que si $p = 1/2$ alors avec probabilité 1, la suite aléatoire $(X_n)_{n\geq 0}$ visite presque sûrement chaque élément de $\mathbb{Z}$ une infinité de fois. En revanche, si $p \neq 1/2$ alors avec probabilité 1, la suite $(X_n)_{n\geq 0}$ ne visite qu'un nombre fini de fois chaque élément de $\mathbb{Z}$. On le voit bien dans les formules du théorème A.15 en faisant tendre a ou b vers l'infini.*

Remarque A.17 (Les théorèmes limites à la rescousse). *Voici un autre argument pour établir que $\mathbb{P}_x(\tau < \infty) = 1$. Posons $m = 2p - 1$ et $\sigma^2 = 4p(1-p)$. Si $m \neq 0$ alors par la loi forte des grands nombres, p.s. $(X_n)_{n\geq 1}$ tend vers $+\infty$ si $m > 0$ et vers $-\infty$ si $m < 0$, et donc $\mathbb{P}_x(\tau < \infty) = 1$. Si $m = 0$ alors pour tout $n \geq 1$, en posant $I_n = \frac{1}{\sqrt{n}}]a, b[$, on a*

$$\mathbb{P}_x(\tau = \infty) \leq \mathbb{P}(a < X_n < b) = \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right).$$

Or $(n^{-1/2}X_n)_{n\geq 1}$ converge en loi vers $\mathcal{N}(0, \sigma^2)$ par le théorème central limite. Mais I_n dépend de n. Cependant, comme $(I_n)_{n\geq 1}$ est décroissante,

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right) \leq \inf_{m\geq 1} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{I_m} e^{-\frac{t^2}{2\sigma^2}}\, dt = 0.$$

Théorème A.18 (Nombres de Catalan). *Si $\tau = \inf\{n \geq 1 : X_n = 0\}$ alors pour tout $n \geq 0$,*

$$\mathbb{P}_0(\tau = 2n+2) = \frac{2}{n+1}\binom{2n}{n} p^{n+1}(1-p)^{n+1}.$$

On reconnaît le n^e nombre de Catalan $\frac{1}{n+1}\binom{2n}{n}$. Ces nombres comptent, outre les chemins de la marche aléatoire simple, les mots de Dyck, les parenthésages, les triangulations d'un polygone, les partitions non croisées, les chemins sous-diagonaux dans le carré, les arbres planaires, etc. Les moments pairs de la loi du demi-cercle sont les nombres de Catalan. C'est l'occasion de souligner que la beauté de la combinatoire réside dans les bijections qu'elle révèle, entre des ensembles finis d'objets de natures a priori très différentes.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Preuve du théorème A.18. Sachant $\{X_0 = 0\}$, l'événement $\{\tau = 2n+2\}$ correspond à une trajectoire de longueur $2n+2$ partant de 0 et revenant à zéro en restant strictement positive ou strictement négative. Ces deux cas sont équiprobables, d'où le facteur 2 dans le résultat. Dans les deux cas, il y a eu forcément $n+1$ incréments $+1$ et $n+1$ incréments -1, d'où

$$\mathbb{P}_0(\tau = 2n+2) = 2C_n p^{n+1}(1-p)^{n+1}.$$

où C_n est le nombre de chemins de longueur $2n+2$ partant de zéro et revenant à zéro, et restant strictement positifs. Le premier incrément est forcément $+1$ et le dernier forcément -1 et C_n est égal au nombre de chemins de longueur $2n$ partant de zéro et revenant à zéro et restant positifs. Il y a n incréments $+1$ et n incréments -1. Considérons les chemins partant de zéro et revenant à zéro et contenant n incréments $+1$ et n incréments -1. Il y en a $\binom{2n}{n}$. Si un chemin de ce type n'est pas positif alors juste après la première position négative, modifions tous les incréments en permutant le signe des $+1$ et des -1. On obtient de la sorte un chemin avec $n-1$ incréments $+1$ et $n+1$ incréments -1, et il s'avère que tous les chemins partant de zéro avec $n-1$ incréments $+1$ et $n+1$ incréments -1 s'obtiennent de la sorte, et il y en a $\binom{2n}{n-1}$. Ainsi, $C_n = \binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}$. Cette astuce est attribuée au mathématicien français Désiré André (1840–1917). □

A.5 Fonctions caractéristiques et vecteurs gaussiens

La loi des variables et vecteur aléatoires ainsi que leur indépendance peuvent être étudiées au moyen de transformées, chacune correspondant à une classe de fonctions test particulière (ces classes sont liées par changement de variable). Pour les variables et vecteurs discrets, c'est la fonction génératrice qu'on a coutume d'utiliser pour sa simplicité. Pour les variables aléatoires positives, c'est plutôt la transformée de Laplace qui est utilisée. Plus généralement, pour des vecteurs aléatoires quelconques, ont utilise la fonction caractéristique, ou transformée de Fourier. La fonction caractéristique des vecteurs aléatoires gaussiens possède des propriétés remarquables. Soulignons que ces transformées sont avant tout liées aux lois, plutôt qu'aux variables ou vecteurs qui suivent ces lois.

Définition A.19 (Fonction caractéristique ou transformée de Fourier). *La fonction caractéristique d'une v.a.r. est la fonction $\Phi_X : \mathbb{R} \to \{z \in \mathbb{C} : |z| \leq 1\}$ définie pour tout $t \in \mathbb{R}$ par $\Phi_X(t) = \mathbb{E}(e^{itX})$. Plus généralement, la fonction caractéristique d'un vecteur aléatoire X de $\mathbb{R}^d$ est la fonction $\Phi_X : \mathbb{R}^d \to \{z \in \mathbb{C} : |z| \leq 1\}$ définie pour tout $t \in \mathbb{R}^d$ par*

$$\Phi_X(t) = \mathbb{E}(e^{i\langle t, X\rangle}).$$

La fonction caractéristique est liée à la fonction génératrice : $g_X(e^{it}) = \Phi_X(t)$. Si X est une v.a.r. telle que X^k est intégrable pour tout $0 \leq k \leq n$ alors Φ_X est n fois dérivable en 0 et $\Phi_X^{(k)}(0) = i^k\mathbb{E}(X^k)$ pour tout $1 \leq k \leq n$. La transformée de Laplace définie par

$$t \in \mathbb{R}^n \mapsto \mathbb{E}(e^{\langle t, X\rangle}) \in \mathbb{R}_+ \cup \{\infty\}$$

n'a pas l'avantage d'être partout finie comme Φ_X. Le calcul effectif de Φ_X peut être mené grâce à la formule du transfert, en utilisant au besoin $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

Théorème A.20 (Caractérisation de la loi – Admis). *Deux vecteurs aléatoires de $\mathbb{R}^d$ ont même loi si et seulement si ils ont même fonction caractéristique.*

Si X est un vecteur aléatoire de $\mathbb{R}^d$ et si $t \in \mathbb{R}$ et $s \in \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ alors $t\langle s, X\rangle = \langle ts, X\rangle$ est une v.a.r. et $\langle s, X\rangle s$ est la projection de X sur la droite $\mathbb{R}s$.

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Corollaire A.21 (Cramér-Wold – Caractérisation par projections). *La loi d'un vecteur aléatoire X de $\mathbb{R}^d$ est caractérisée par les l'ensemble des lois de $\langle s, X\rangle$ pour tout $s \in \mathbb{S}^{d-1}$.*

Corollaire A.22 (Caractérisation de l'indépendance). *Deux vecteurs aléatoires X et Y de $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ sont indépendants si et seulement si pour tous $s \in \mathbb{R}^d$ et $t \in \mathbb{R}^d$*

$$\Phi_{(X,Y)}(s,t) = \Phi_X(s)\Phi_Y(t).$$

En particulier, si X et Y sont indépendants et $d = d'$ alors pour tout $t \in \mathbb{R}^d$,

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

Soit X une v.a.r. telle que X^n est intégrable pour tout $n \in \mathbb{N}$. On dit que la loi de X est caractérisée par ses moments lorsque pour toute v.a.r. Y, si on a $\mathbb{E}(Y^n) = \mathbb{E}(X^n)$ pour tout $n \in \mathbb{N}$ alors la v.a.r. Y a la même loi que X. Le résultat suivant entraîne que la loi normale standard $\mathcal{N}(0,1)$ et la loi exponentielle sont caractérisées par leurs moments.

Théorème A.23 (Théorème des moments de Stieltjes). *Soit X une v.a.r. avec X^n intégrable pour tout $n \in \mathbb{N}$. Posons $m_n = \mathbb{E}(X^n)$. Les propositions suivantes sont équivalentes :*

1. *Φ_X est analytique sur un voisinage de 0*
2. *Φ_X est analytique sur $\mathbb{R}$*
3. *$\overline{\lim}_{n\to\infty} \left(\frac{1}{n!}|m_n|\right)^{\frac{1}{n}} < \infty$.*

Si ces conditions sont vérifiées alors la loi de X est caractérisée par ses moments.
En particulier, une loi à support compact est caractérisée par ses moments.

La formule de Stirling donne $(1/n!)^{1/n} = \mathcal{O}_{n\to\infty}(1/n)$. Par conséquent, la loi de X est caractérisée par ses moments si $\overline{\lim}_{n\to\infty} \frac{1}{n}|m_n|^{1/n} < \infty$.

Démonstration. Pour tout $n \in \mathbb{N}$, on a $\mathbb{E}(|X|^n) < \infty$ et donc Φ_X est n fois dérivable sur $\mathbb{R}$. De plus, $\Phi_X^{(n)}$ est continue sur $\mathbb{R}$ et pour tout $t \in \mathbb{R}$,

$$\Phi_X^{(n)}(t) = \mathbb{E}((iX)^n e^{itX}).$$

En particulier, $\Phi_X^{(n)}(0) = i^n m_n$, et la série de Taylor de Φ_X en 0 est déterminée par la suite $(m_n)_{n\geq 1}$. Comme le rayon de convergence r d'une série entière $\sum_n a_n z^n$ est donné par la formule de Hadamard $r^{-1} = \overline{\lim}_n |a_n|^{\frac{1}{n}}$, on obtient que 1 $\Rightarrow$ 3 (prendre $a_n = i^n m_n/n!$). D'autre part, comme pour tout $n \in \mathbb{N}$ et tous $s, t \in \mathbb{R}$,

$$\left|e^{isx}\left(e^{itx} - 1 - \frac{itx}{1!} - \cdots - \frac{(itx)^{n-1}}{(n-1)!}\right)\right| \leq \frac{|tx|^n}{n!},$$

on a pour tout $n \in \mathbb{N}$ pair et tous $s, t \in \mathbb{R}$,

$$\left|\Phi_X(s+t) - \Phi_X(s) - \frac{t}{1!}\Phi_X'(s) - \cdots - \frac{t^{n-1}}{(n-1)!}\Phi_X^{(n-1)}(s)\right| \leq m_n \frac{|t|^n}{n!},$$

qui montre que 3 $\Rightarrow$ 2. Comme 2 $\Rightarrow$ 1 on a bien équivalence de 1 2 3. Si X est bornée, alors $\sup_n |m_n|^{\frac{1}{n}} < \infty$ et donc 3 a lieu en utilisant la formule de Stirling. Si 3 a lieu alors les arguments précédents donnent un $r > 0$ tel que Φ_X est développable en série entière en tout $x \in \mathbb{R}$ avec un rayon de convergence $\geq r$. De proche en proche, on obtient que Φ_X est caractérisée par ses dérivées en zéro. □

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Théorème A.24 (Vecteurs et lois gaussiennes). *Si X est un vecteur aléatoire de $\mathbb{R}^d$ de moyenne m et de matrice de covariance Σ alors les propriétés suivantes sont équivalentes :*

1. *Toute combinaison linéaire des composantes de X suit une loi normale sur $\mathbb{R}$*
2. *La fonction caractéristique de X est donnée pour tout $t \in \mathbb{R}^d$ par*

$$\Phi_X(t) = \mathbb{E}\Big(e^{i\langle t,X\rangle}\Big) = \exp\left(i\langle t,m\rangle - \frac{1}{2}\langle t,\Sigma t\rangle\right).$$

3. *$\mathcal{L}(X) = \mathcal{L}(m + AZ)$, où A est une matrice de dimension $d \times d$ vérifiant $AA^\top = \Sigma$ et Z est un vecteur aléatoire de $\mathbb{R}^d$ à composantes indépendantes et de loi $\mathcal{N}(0,1)$.*

On dit alors que X est un *vecteur gaussien*. Sa loi est caractérisée par son vecteur moyenne m et sa matrice de covariance Σ. Elle est notée $\mathcal{N}(m,\Sigma)$. On dit que c'est une loi gaussienne sur $\mathbb{R}^d$. La loi $\mathcal{N}(0, I_d)$ de Z est appelée *loi gaussienne standard*.

Démonstration. Une combinaison linéaire des composantes de X s'écrit $\langle u, X\rangle = u^\top X$ où u est un vecteur colonne déterministe de $\mathbb{R}^d$. L'équivalence 1) $\Leftrightarrow$ 2) découle donc de l'expression de la fonction caractéristique des lois gaussiennes sur $\mathbb{R}$, et du fait que la fonction caractéristique caractérise la loi. Cette dernière propriété montre également que la loi de X est caractérisée par m et Σ. L'équivalence 1) $\Leftrightarrow$ 3) provient du théorème 4.24 sur la racine carrée matricielle, associé au théorème 4.25 sur la transformation linéaire. □

Exemple A.25 (Nécessaire mais pas suffisant). *Les composantes d'un vecteur gaussien sont gaussiennes, mais la réciproque est fausse. En effet, soit $X = (Y, \varepsilon Y)$ un vecteur aléatoire de $\mathbb{R}^2$ où Y et ε sont indépendantes avec $Y \sim \mathcal{N}(0,1)$ sur $\mathbb{R}$ et ε de loi de Rademacher symétrique : $\mathbb{P}(\varepsilon = \pm 1) = 1/2$. Les composantes Y et εY de X sont gaussiennes mais la combinaison linéaire $Y + \varepsilon Y$ ne l'est pas car $\mathbb{P}(Y + \varepsilon Y = 0) = \mathbb{P}(\varepsilon = -1) = 1/2$. De plus, $\mathrm{Cov}(Y, \varepsilon Y) = \mathbb{E}(Y^2)\mathbb{E}(\varepsilon) = 0$ mais Y et εY ne sont pas indépendantes.*

Théorème A.26 (Existence de densité). *La loi gaussienne $\mathcal{N}(m,\Sigma)$ sur $\mathbb{R}^d$ admet une densité de probabilité si et seulement si Σ est inversible donnée dans ce cas par*

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\Big(-\frac{1}{2}\langle (x-m), \Sigma^{-1}(x-m)\rangle\Big).$$

Démonstration. Soit A une racine carrée matricielle de Σ de même rang $p \leq d$ que Σ. Soit Z un vecteur gaussien standard Z de loi $\mathcal{N}(0, \mathrm{I}_d)$, de sorte que $m + AZ \sim \mathcal{N}(m,\Sigma)$. La loi $\mathcal{N}(m,\Sigma)$ est portée par le sous-espace affine $E = \{Az + m$ avec $z \in \mathbb{R}^d\}$ de dimension p. Si $p < d$, alors $E \neq \mathbb{R}^d$ et $\mathcal{N}(m,\Sigma)$ n'a pas de densité. De plus, $E = \mathbb{R}^d$ si et seulement si $p = d$, c'est-à-dire si et seulement si Σ est inversible. On peut choisir les composantes de Z indépendantes et de même loi $\mathcal{N}(0,1)$ sur $\mathbb{R}$. La loi de Z admet alors la densité de probabilité f_Z donnée, pour tout $z \in \mathbb{R}^d$, par

$$f_Z(z) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_k^2\right) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\|z\|_2^2\right).$$

Si $X \sim \mathcal{N}(m,\Sigma)$, alors pour toute indicatrice de pavé $h : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}(h(X)) = \mathbb{E}(h(AZ + m)) = \int_{\mathbb{R}^d} h(Az + m) f_Z(z)\, dz.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Si Σ est inversible, le changement de variable affine $x = Az + m$ est un difféomorphisme de $\mathbb{R}^d$ dans lui-même, de jacobien non nul égal à $\det(A^{-1})$. La décomposition $\Sigma = AA^\top$ entraîne que $|\det(A)| = \sqrt{\det(\Sigma)}$ et $\Sigma^{-1} = (AA^\top)^{-1} = (A^{-1})^\top A^{-1}$. On en déduit que

$$\mathbb{E}[h(X)] = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \int_{\mathbb{R}^d} h(x) \exp\Big(-\frac{1}{2}(x-m)^\top \Sigma^{-1}(x-m)\Big)\, dx$$

d'où la formule annoncée pour la densité f. □

Théorème A.27 (Indépendance des composantes). *Pour tout vecteur gaussien X de $\mathbb{R}^d$, les trois propriétés suivantes sont équivalentes :*

1. *Les composantes $X_1, \ldots, X_d$ sont mutuellement indépendantes*
2. *Les composantes $X_1, \ldots, X_d$ sont deux à deux indépendantes*
3. *La matrice de covariance Σ de X est diagonale.*

En particulier, un vecteur aléatoire gaussien est gaussien standard si et seulement si ses composantes sont indépendantes et de même loi normale centrée réduite $\mathcal{N}(0,1)$ sur $\mathbb{R}$.

Démonstration. Les implications 1) ⇒ 2) et 2) ⇒ 3) découlent des définitions. Vérifions que 3) ⇒ 1). Si on a $\Sigma = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_d^2)$, alors pour tout $t \in \mathbb{R}^d$,

$$\Phi_X(t) = \exp\Big(i\langle t, m\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\Big) = \prod_{k=1}^d \exp\Big(it_k m_k - \frac{1}{2}\sigma_k^2 t_k^2\Big) = \prod_{k=1}^d \Phi_{X_k}(t_k).$$

□

Les lois gaussiennes sont stables par transformation affine. En effet, si $X \sim \mathcal{N}(m, \Sigma)$ et $A \in \mathcal{M}_{p,d}(\mathbb{R})$ et $b \in \mathbb{R}^p$ alors $AX + b \sim \mathcal{N}(Am + b, A\Sigma A^\top)$. En particulier, $\mathcal{N}(0, \sigma^2 \mathrm{I}_d)$ est invariante par rotation et symétries car si $X \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$ et si A est $d \times d$ orthogonale alors AX a la même loi que X. Le théorème de Cochran raffine l'étude de $\mathcal{N}(0, \sigma^2 \mathrm{I}_d)$.

Théorème A.28 (Cochran). *Soit X un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(m, \sigma^2 \mathrm{I}_n)$ et $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_p$ une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions $d_1, \ldots, d_p$ avec $d_1 + \cdots + d_p = n$. Soit $\mathbf{P}_k$ la matrice du projecteur orthogonal sur E_k et $Y_k = \mathbf{P}_k X$ la projection orthogonale de X sur E_k.*

1. *Les projections $(Y_1, \ldots, Y_p)$ sont des vecteurs gaussiens indépendants et*

$$Y_k \sim \mathcal{N}(\mathbf{P}_k m, \sigma^2 \mathbf{P}_k).$$

2. *Les variables aléatoires $\|Y_1 - \mathbf{P}_1 m\|_2^2, \ldots, \|Y_p - \mathbf{P}_p m\|_2^2$ sont indépendantes et*

$$\sigma^{-2}\|Y_k - \mathbf{P}_k m\|_2^2 \sim \chi^2(d_k).$$

Démonstration. On se ramène d'abord au cas où $m = 0$ par translation. Le vecteur aléatoire $Y = (Y_1, \ldots, Y_p)^\top$ de $\mathbb{R}^{np}$ s'écrit $Y = AX$ où A est la matrice de dimension $np \times n$

$$A = \begin{pmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_p \end{pmatrix}.$$

Il en découle que Y suit la loi $\mathcal{N}(0, \sigma^2 AA^\top)$. Pour tout $1 \le i \le p$, on a $\mathbf{P}_i = \mathbf{P}_i^\top = \mathbf{P}_i^2$. De plus, $\mathbf{P}_i\mathbf{P}_j = 0$ si $1 \le i \ne j \le p$ car $E_i \perp E_j$. Par conséquent, $AA^\top = \mathrm{Diag}(\mathbf{P}_1, \ldots, \mathbf{P}_p)$ est diagonale par blocs. On peut déduire du théorème A.27 que $Y_1, \ldots, Y_p$ sont des vecteurs

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

gaussiens indépendants avec $Y_k \sim \mathcal{N}(0, \sigma^2 \mathbf{P}_k)$ pour tout $1 \leq k \leq p$. En particulier, les variables aléatoires $\|Y_1\|_2^2, \ldots, \|Y_p\|_2^2$ sont indépendantes. Il reste à déterminer leur loi. Pour tout $1 \leq k \leq p$, soit $B_k = \{e_{k,1}, \ldots, e_{k,d_k}\}$ une base orthonormée de E_k. La réunion $B_1 \cup \cdots \cup B_p$ constitue une base orthonormée de $\mathbb{R}^n$. Le vecteur X s'écrit dans cette base $X = Y_1 + \cdots + Y_p$ avec $Y_k = a_{k,1} e_{k,1} + \cdots + a_{k,d_k} e_{k,d_k}$ où $a_{k,i} = \langle X, e_{k,i} \rangle$. L'invariance par transformation orthogonale de la loi $\mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ implique que les variables aléatoires $a_{k,i}$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$. Il en découle que pour tout $1 \leq k \leq p$,

$$\sigma^{-2} \|Y_k\|^2 = \sigma^{-2}(a_{k,1}^2 + \cdots + a_{k,d_k}^2) \sim \chi^2(d_k).$$

□

Corollaire A.29 (Échantillons gaussiens). *Soient $X_1, \ldots, X_n$ des v.a.r. de loi normale $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$, de moyenne empirique et variance empirique définies par*

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \qquad \text{et} \qquad S_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2.$$

Alors les variables aléatoires $\overline{X}_n$ et S_n^2 sont indépendantes avec

$$\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \qquad \text{et} \qquad \frac{(n-1)}{\sigma^2} S_n^2 \sim \chi^2(n-1).$$

De plus, la moyenne empirique studentisée T_n vérifie

$$T_n = \sqrt{n} \left(\frac{\overline{X}_n - m}{S_n} \right) \sim t(n-1).$$

Démonstration. Soit 1_n le vecteur de $\mathbb{R}^n$ dont toutes les coordonnées sont égales à 1. La matrice de la projection orthogonale sur $E_1 = \mathbb{R} 1_n$ est donnée par

$$\mathbf{P}_1 = \frac{1_n 1_n^\top}{\|1_n\|^2} = \frac{1}{n} 1_n 1_n^\top.$$

Le sous-espace $E_2 = E_1^\perp$ est de dimension $n-1$ et la matrice de la projection orthogonale sur E_2 est $\mathbf{P}_2 = 1_n - \mathbf{P}_1$. On a $Y_1 = \mathbf{P}_1 X = \overline{X}_n 1_n$ et $Y_2 = \mathbf{P}_2 X = (X_1 - \overline{X}_n, \ldots, X_n - \overline{X}_n)^\top$, ce qui entraîne $\|Y_2\|^2 = (n-1) S_n^2$. Le théorème de Cochran permet de conclure. □

Théorème A.30 (Paul Lévy – Admis). *Si $(X_n)_{n \geq 1}$ et X sont des vecteurs aléatoires de $\mathbb{R}^d$ et si X admet une densité continue alors les propriétés suivantes sont équivalentes :*

1. $\lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ *pour toute fonction continue et bornée* $f : \mathbb{R}^d \to \mathbb{R}$
2. $\lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ *pour toute indicatrice f de pavé ou de boule*
3. $\lim_{n \to \infty} \varphi_{X_n}(t) = \varphi_X(t)$ *pour tout* $t \in \mathbb{R}^d$

Voici une version multivariée du théorème de la limite centrale 5.12.

Corollaire A.31 (Théorème de la limite centrale multivarié). *Soit $(X_n)_{n \geq 1}$ une suite de vecteurs aléatoires de $\mathbb{R}^d$ indépendants et de même loi, dont les composantes sont de carré intégrable. Alors, en notant m et Σ le vecteur moyenne et la matrice de covariance de X_1, on a, avec $X \sim \mathcal{N}(0, \Sigma)$, pour tout pavé ou boule B de $\mathbb{R}^d$,*

$$\lim_{n \to \infty} \mathbb{P}\left(\sqrt{n} \left(\frac{X_1 + \cdots + X_n}{n} - m \right) \in B \right) = \mathbb{P}(X \in B).$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. Le théorème de Paul Lévy ramène le problème à la convergence ponctuelle des fonctions caractéristiques vers celle de la loi gaussienne $\mathcal{N}(0,\Sigma)$. Quitte à remplacer les X_k par $X_k - m$, on peut supposer que $m = 0$. Comme $X_1,\dots,X_n$ sont des vecteurs aléatoires indépendants et de même loi, on a pour tout $t \in \mathbb{R}^d$

$$\Phi_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(t) = \mathbb{E}\left(\exp\left(i\left\langle \frac{t}{\sqrt{n}}, X_1\right\rangle + \cdots + i\left\langle \frac{t}{\sqrt{n}}, X_n\right\rangle\right)\right) = \left(\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Pour tout $t \in \mathbb{R}^n$, la v.a.r. $\langle t, X_1\rangle$ a pour moyenne 0 et pour variance $\langle t, \Sigma t\rangle$. Une formule de Taylor à l'ordre 2 en 0 pour $\Phi_{\langle t, X_1\rangle}$ donne $\Phi_{X_1}(t) = 1 + \frac{1}{2}\langle t, \Sigma t\rangle + o_{t\to 0}(\|t\|_2^2)$ d'où

$$\Phi_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(t) = \left(1 + \frac{1}{2n}\langle t, \Sigma t\rangle + \|t\|_2^2 o_{n\to\infty}\left(\frac{1}{n}\right)\right)^n \to \exp\left(\frac{1}{2}\langle t, \Sigma t\rangle\right) = \Phi_{\mathcal{N}(0,\Sigma)}(t).$$

□

Bien qu'il ne s'agisse pas d'une véritable distance, on appelle distance du χ^2 entre deux lois de probabilité p et q sur un ensemble fini $\{1,\dots,k\}$ le nombre réel positif

$$D(p,q) = \sum_{i=1}^{k} \frac{(p_i - q_i)^2}{p_i}.$$

Cette quantité asymétrique en p et q vaut $+\infty$ si l'un des p_i est nul.

Théorème A.32 (Test d'adéquation du χ^2). *Soit $p = (p_1,\dots,p_k)$ une loi sur $\{1,\dots,k\}$ et $X_1,\dots,X_n$ des v.a.r. sur $\{1,\dots,k\}$ indépendantes et de loi $q = (q_1,\dots,q_k)$. On définit les effectifs théoriques $n_1,\dots,n_k$, empiriques $N_1,\dots,N_k$, et la loi $\widehat{p} = (\widehat{p}_1,\dots,\widehat{p}_k)$ par*

$$n_i = np_i \quad \text{et} \quad N_i = \mathbf{1}_{\{X_1=i\}} + \cdots + \mathbf{1}_{\{X_n=i\}} \quad \text{et} \quad \widehat{p}_i = \frac{N_i}{n}$$

Supposons que $p_1 > 0,\dots,p_k > 0$. Considérons la distance du χ^2 normalisée

$$D_n = nD(p,\widehat{p}) = n\sum_{i=1}^{k} \frac{(p_i - \widehat{p}_i)^2}{p_i} = \sum_{i=1}^{k} \frac{(n_i - N_i)^2}{n_i}.$$

Si $p = q$ alors pour tout $t \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{D_n}(t) = F_{\chi^2(k-1)}(t)$$

Si $p \neq q$ alors avec probabilité 1,

$$\lim_{n\to\infty} D_n = +\infty.$$

Démonstration. Supposons que $p \neq q$. La loi forte des grands nombres du théorème 5.4 entraîne qu'avec probabilité 1, on a $\lim_{n\to\infty} N_i/n \to q_i$ pour tout $1 \le i \le k$ et donc

$$\lim_{n\to\infty} \frac{D_n}{n} = \sum_{i=1}^{k} \frac{(p_i - q_i)^2}{p_i} = D(p,q)$$

et comme $D(p,q) > 0$ car $p \neq q$, on obtient bien $\lim_{n\to\infty} D_n = +\infty$. Supposons à présent au contraire que $p = q$. Pour $1 \le j \le n$, soit V_j le vecteur aléatoire de $\mathbb{R}^k$ donné par

$$(V_j)_i = \frac{1}{\sqrt{p_i}}(\mathbf{1}_{\{X_i=i\}} - p_i).$$

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Les vecteurs $V_1, \ldots, V_n$ sont indépendantes et de même loi, et cette loi est centrée et de matrice de covariance $\Sigma = I_k - \sqrt{p}\sqrt{p}^\top$ avec $\sqrt{p}^\top = (\sqrt{p_1}, \ldots, \sqrt{p_k})$. Le théorème A.31 de la limite centrale multivarié entraîne alors que pour tout pavé ou boule B de $\mathbb{R}^k$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}}(V_1 + \cdots + V_n) \in B\right) = \mathbb{P}(Z \in B)$$

où $Z \sim \mathcal{N}(0, \Sigma)$. Soit $\mathrm{Vect}(\sqrt{p})$ le sous-espace de $\mathbb{R}^k$ engendré par $\sqrt{p}$ et soit $H = \sqrt{p}\sqrt{p}^\top$ la matrice de projection orthogonale sur $\mathrm{Vect}(\sqrt{p})$. La matrice de projection orthogonale sur $\mathrm{Vect}(\sqrt{p})^\perp$ est $I_k - H = \Sigma$. Cette matrice est de rang $k-1$ car H est de rang 1, et le théorème A.28 de Cochran donne $\|Z\|_2^2 \sim \chi^2(k-1)$. Il ne reste plus qu'à observer que

$$D_n = \left\|\frac{1}{\sqrt{n}}(V_1 + \cdots + V_n)\right\|_2^2.$$

□

Dans la pratique, on connaît p mais pas q, et on souhaite décider au vu de $X_1, \ldots, X_n$ si $p = q$ ou non. Cette décision est prise au moyen d'un test d'adéquation asymptotique. Plus précisément, on fixe un niveau de confiance $\alpha \in (0,1)$ comme par exemple $\alpha = 0{,}05$, puis on détermine le quantile a_α d'ordre $1-\alpha$ de la loi du chi-deux $\chi^2(k-1)$, ce qui donne la région d'acceptation du test $\mathcal{A}_\alpha = [0, a_\alpha]$. La règle de décision est la suivante :

si $D_n \in \mathcal{A}_\alpha$ alors on accepte l'hypothèse $p = q$ et sinon on la rejette.

La probabilité de rejeter à tort tend vers α quand $n \to \infty$ (erreur de première espèce).
La probabilité d'accepter à tort tend vers 0 quand $n \to \infty$ (erreur de seconde espèce).
À $X_1, \ldots, X_n$ fixés, plus α est petit, moins on rejette à tort mais plus on accepte à tort.

A.6 Extrêmes

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi et

$$M_n = \max(X_1, \ldots, X_n).$$

Pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}$ on a

$$F_{M_n}(x) = \mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x) \cdots \mathbb{P}(X_n \leq x) = F_{X_1}^n(x).$$

Cette quantité tend vers 0 ou 1 selon que $F_{X_1}(x) < 1$ ou $F_{X_1}(x) = 1$. Cela suggère que le comportement asymptotique de M_n dépend du comportement de la loi de X_1 au bord droit de son support. On cherche donc $(a_n, b_n)_{n\geq 1}$ de sorte que $\mathbb{P}(a_n M_n + b_n \leq x)$ tende vers une fonction de répartition d'une loi, si possible non dégénérée c'est-à-dire non Dirac. Cela n'est pas toujours possible : si par exemple X_1 suit la loi de Bernoulli $\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0) = p$ avec $0 < p < 1$ alors $M_n = 1$ sur $\{n \geq T\}$ où $T = \inf\{n \geq 1 : X_n = 1\}$, et comme T suit une loi géométrique de moyenne $1/p$, on a $\mathbb{P}(T < \infty) = 1$, d'où $\mathbb{P}(M_n \to 1) = 1$. Plus généralement, cela est impossible si la loi de X_1 possède un atome au bord droit de son support, comme le montre le résultat suivant.

Théorème A.33 (Bord droit du support). *On a toujours*

$$\mathbb{P}\left(\lim_{n\to\infty} M_n = x_F\right) = 1 \quad \text{où} \quad x_F = \sup\{x \in \mathbb{R} : F_{X_1}(x) < 1\} \in \mathbb{R} \cup \{\infty\}.$$

Copyright © Djalil Chafaï 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material

Démonstration. Pour tout $x < x_F$, on a $F_{X_1}(x) < 1$ et donc

$$\mathbb{P}(M_n \leq x) = F_{X_1}^n(x) \underset{n\to+\infty}{\longrightarrow} 0.$$

De plus, dans le cas où $x_F < \infty$, on a pour tout $x \geq x_F$, $F(x) = 1$, et donc

$$\mathbb{P}(M_n \leq x) = F_{X_1}^n(x) \to 1.$$

Ainsi, la suite $(M_n)_{n\geq 1}$ converge en probabilité vers x_F, et comme elle est croissante, elle converge presque sûrement vers la même limite x_F. □

Examinons trois situations dont le comportement est facile à étudier : X_1 est uniforme (queue à droite nulle), X_1 est exponentielle (queue à droite à décroissance exponentielle), X_1 est de Cauchy (queue à droite à décroissante polynomiale).

Théorème A.34 (Unif. $\rightsquigarrow$ Weibull). *Si X_1 suit la loi uniforme sur $[0,\theta]$ avec $\theta > 0$ alors*

$$\lim_{n\to\infty} F_{n(\theta^{-1}M_n-1)}(x) = e^x \mathbf{1}_{\mathbb{R}_-}(x) + \mathbf{1}_{\mathbb{R}_+}(x).$$

pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition de $-E$ où E suit une loi exponentielle (on dit parfois qu'il s'agit de la loi de Weibull des extrêmes, à ne pas confondre avec les lois de Weibull utilisées pour modéliser les durées de vie).

Démonstration. Pour tout $x \leq 0$ on a $n^{-1}x + 1 \leq 1$ et

$$\mathbb{P}(M_n \leq \theta(n^{-1}x + 1)) = (n^{-1}x + 1)^n \to e^x.$$

Remarque : cela donne la vitesse et la loi de fluctuation (non gaussienne) de l'estimateur $\widehat{\theta}_n = \max\{U_1, \ldots, U_n\}$ de θ où $U_1, \ldots, U_n$ sont i.i.d. de loi uniforme sur $[0,\theta]$ (il s'agit d'un modèle statistique non régulier). □

Le théorème (A.34) indique une fluctuation non normale pour l'estimateur M_n de θ.

Théorème A.35 (Expo. $\rightsquigarrow$ Gumbel). *Si X_1 est exponentielle de moyenne $1/\lambda$ alors*

$$\lim_{n\to\infty} F_{\lambda M_n - \ln(n)}(x) = e^{-e^{-x}}$$

pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition d'une loi de Gumbel.

Application : on peut approcher $\mathbb{P}(M_n \leq x)$ par $F(\lambda x - \ln(n))$ où F est la fonction de répartition de la loi de Gumbel. Par exemple pour $\lambda = 1/10$, $x = 50$, $n = 100$, on a

$$\mathbb{P}(M_{100} \geq 50) \approx e^{-e^{-(\lambda x - \ln(n))}} = 0,49023$$

tandis que le calcul exact donne

$$\mathbb{P}(M_{100} \geq 50) = 1 - (1 - e^{-\lambda x})^n = 0,49139.$$

Démonstration. Pour tout $x \in \mathbb{R}$ tel que $x + \ln(n) \geq 0$ (toujours vrai si n assez grand)

$$\mathbb{P}(\lambda M_n - \ln(n) \leq x) = (1 - n^{-1}e^{-x})^n \to e^{-e^{-x}}.$$

□

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Théorème A.36 (Cauchy $\rightsquigarrow$ Fréchet). *Si X_1 suit la loi de Cauchy alors*

$$\lim_{n\to\infty} F_{\pi n^{-1} M_n}(x) = e^{-1/x}\mathbf{1}_{\mathbb{R}_+^*}(x)$$

pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition de la loi de Fréchet.

Démonstration. Comme $\arctan(x) = \pi/2 - 1/x + \mathcal{O}_{x\to\infty}(1/x^2)$, pour tout $x \geq 0$,

$$\mathbb{P}\Big(M_n \leq \frac{nx}{\pi}\Big) = \left(\int_{-\infty}^{nx\pi^{-1}} \frac{dy}{\pi(1+x^2)}\right)^n = \left(1 - \frac{1}{nx} + \mathcal{O}_{x\to\infty}(n^{-2})\right)^n \to e^{-1/x}.$$

□

Ainsi, lorsque X_1 suit une loi uniforme, une loi exponentielle, ou une loi de Cauchy, il existe une suite déterministe $(a_n, b_n)_{n\geq 1}$ avec $a_n > 0$ ainsi qu'une loi L sur $\mathbb{R}$ de fonction de répartition F tel que pour tout $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{a_n M_n + b_n}(x) = F(x).$$

Le théorème suivant affirme que quelque soit la loi de X_1, la loi L est soit une masse de Dirac soit l'une des trois lois déjà observées à translation et dilatation près.

Théorème A.37 (Extrêmes de Gnedenko-Fréchet-Fisher-Tippett). *S'il existe une suite $(a_n, b_n)_{n\geq 1}$ et une loi L de fonction de répartition F telles que pour tout $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} F_{a_n M_n + b_n}(x) = F(x)$$

alors, soit L est une Dirac, soit L est à translation/dilatation près...

1. *une loi de Weibull*[1] *avec $F(x) = e^{-(-x)^\alpha}\mathbf{1}_{\mathbb{R}_-}(x) + \mathbf{1}_{\mathbb{R}_+}(x)$ pour un paramètre $\alpha > 0$*
2. *une loi de Gumbel avec $F(x) = e^{-e^{-x}}\mathbf{1}_{\mathbb{R}}(x)$*
3. *une loi de Fréchet avec $F(x) = e^{-x^{-\alpha}}\mathbf{1}_{\mathbb{R}_+}(x)$ pour un paramètre $\alpha > 0$*

La théorie des extrêmes fournit des conditions nécessaires et suffisantes sur F_{X_1} pour l'appartenance aux bassins d'attraction des trois lois des extrêmes. Ces conditions portent sur la queue à droite. La loi de Weibull apparaît pour les lois dont la queue à droite est nulle (ex. loi uniforme), la loi de Gumbel apparaît pour les lois dont la queue à droite est exponentielle (ex. lois exponentielle et normale), et la loi de Fréchet apparaît pour les lois dont la queue à droite est polynomiale (lois de Cauchy, de Student, de Pareto).

Remarque A.38 (Lois des extrêmes).

1. *Si X est Weibull de paramètre α alors $M_n \stackrel{d}{=} n^{-1/\alpha}X_1$*
2. *Si X est Gumbel alors $M_n \stackrel{d}{=} X_1 + \ln(n)$*
3. *Si X est Fréchet de paramètre α alors $M_n \stackrel{d}{=} n^{1/\alpha}X_1$*

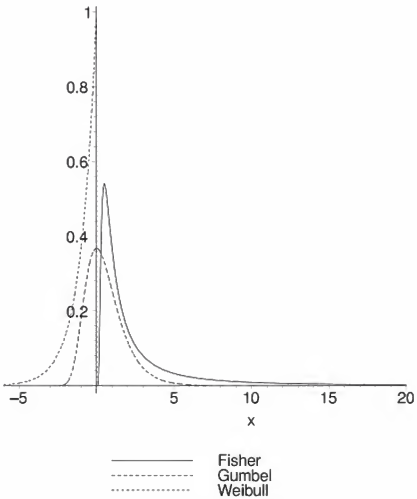
En particulier, ces trois lois sont max-stables : $M_n \stackrel{d}{=} a_n X_1 + b_n$ pour un (a_n, b_n).
De plus, pour tout $\alpha > 0$ et toute variable aléatoire X, il y a équivalence entre :

1. *$-X^{-1}$ suit une loi de Weibull de paramètre α*
2. *$\ln(X^\alpha)$ suit une loi de Gumbel*
3. *X suit une loi de Fréchet de paramètre α*

1. Attention, la terminologie diffère de celle utilisée pour modéliser les durées de vie en fiabilité/survie.

Copyright © Djalil Chafai 2012. ISBN 978-2-9541710-0-5 EAN 9782954171005

Copyrighted material



Copyright © Djalil Chafai 2012 ISBN 978-2-9541710-0-5 EAN 9782954171005
Copyrighted material

Annexe B

Leçons d'oral

Leçons d'analyse et probabilités

229 : Suite de variables aléatoires indépendantes de même loi de Bernoulli. Variable aléatoire de loi binomiale. Approximations de cette loi.

230 : Probabilité conditionnelle et indépendance. Variables aléatoires indépendantes. Variance, covariance.

231 : Espérance, variance ; loi faible des grands nombres.

232 : Variables aléatoires possédant une densité. Exemples.

249 : Loi normale en probabilités.

Exemples et exercices d'analyse et probabilités

435 : Exemples d'étude probabiliste de situations concrètes

437 : Exercices faisant intervenir des variables aléatoires

442 : Exemples d'exercices faisant intervenir le calcul des probabilités

Copyrighted material